



Variational data assimilation in the land surface model ORCHIDEE using YAO

Hector Simon Benavides Pinjosovsky

► To cite this version:

Hector Simon Benavides Pinjosovsky. Variational data assimilation in the land surface model ORCHIDEE using YAO. Earth Sciences. Université Pierre et Marie Curie - Paris VI, 2014. English. NNT : 2014PA066590 . tel-01145923

HAL Id: tel-01145923

<https://theses.hal.science/tel-01145923>

Submitted on 27 Apr 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université Pierre et Marie Curie

Ecole doctorale de Sciences de l'environnement de l'Ile de France

Spécialité

Méthodes statistiques et modèles inverses

Sujet de la thèse

Assimilation variationnelle des données dans le modèle de surface
continentale ORCHIDEE grâce au logiciel YAO

Présentée par

Hector Simon Benavides Pinjosovsky

Dirigée par

Sylvie Thiria et Catherine Ottlé

Présentée et soutenue publiquement le 27 mars 2014

Devant le jury composé de

Mme. Sylvie Thiria	LOCEAN/IPSL	Directeur de thèse
Mme. Catherine Ottlé	LSCE/CEA/IPSL	Co-directeur de thèse
Mme. Isabelle Herlin	INRIA	Rapporteur
Mme. Dominique Courault	INRA	Rapporteur
M. Fouad Badran	CEDRIC/CNAM	Examineur
M. Hervé Le Treut	LMD/CNRS	Examineur
M. Jacques Zegbeu Poussi	CLIMMOD	Examineur

Résumé

Un modèle de surface continentale (LSM en anglais) est un modèle numérique décrivant les échanges d'eau et d'énergie entre la surface terrestre et l'atmosphère. La physique de la surface de la terre comprend une vaste collection de processus complexes. L'équilibre entre la complexité du modèle et sa résolution, confronté à des limitations de calcul, représente une question fondamentale dans le développement d'un LSM. Les observations des phénomènes étudiés sont nécessaires afin d'adapter la valeur des paramètres du modèle à des variables reproduisant le monde réel. Le processus d'étalonnage consiste en une recherche des paramètres du modèle qui minimisent l'écart entre les résultats du modèle et un ensemble d'observations. Dans ce travail, nous montrons comment l'assimilation variationnelle de données est appliquée aux bilans d'énergie et d'eau du modèle de surface continentale ORCHIDEE afin d'étalonner les paramètres internes du modèle. Cette partie du modèle est appelé SECHIBA. Le logiciel YAO est utilisé pour faciliter la mise en œuvre de l'assimilation variationnelle 4DVAR.

Une analyse de sensibilité a été réalisée afin d'identifier les paramètres les plus influents sur la température. Avec la hiérarchie des paramètres obtenue, des expériences jumelles à partir d'observations synthétiques ont été mises en œuvre. Les résultats obtenus suggèrent que l'assimilation de la température de surface a le potentiel d'améliorer les estimations de variables, en ajustant correctement les paramètres de contrôle. Enfin, plusieurs assimilations ont été faites en utilisant des observations de données réelles du site SMOSREX à Toulouse, France. Les expériences faites en utilisant différentes valeurs initiales pour les paramètres, montrent les limites de l'assimilation de la température pour contraindre les paramètres de contrôle. Même si l'estimation des variables est améliorée, ceci est dû à des valeurs finales des paramètres aux limites des intervalles prescrit de la fonction de coût. Afin de parvenir à un minimum, il faudrait permettre aux paramètres de visiter des valeurs irréalistes. Les résultats montrent que SECHIBA ne simule pas correctement simultanément la température et les flux et la relation entre les deux n'est pas toujours cohérente selon le régime (ou les valeurs des paramètres que l'on utilise). Il faut donc travailler sur la physique pour mieux simuler la température. En outre, la sensibilité des paramètres à la température n'est pas toujours suffisante, donnant une fonction de coût plate dans l'espace des paramètres prescrit. Nos résultats montrent que le système d'assimilation mis en place est robuste, puisque les résultats des expériences jumelles sont satisfaisants.

Le couplage entre l'hydrologie et la thermodynamique dans SECHIBA doit donc être revu afin d'améliorer l'estimation des variables. Une étude exhaustive de l'erreur des mesures doit être menée afin de récupérer des termes de pondération dans la fonction de coût. Enfin, l'assimilation d'autres variables telles que l'humidité du sol peut maintenant être réalisée afin d'évaluer l'impact sur les performances de l'assimilation.

Abstract

A land surface model (LSM) is a numerical model describing the exchange of water and energy between the land surface and the atmosphere. Land surface physics includes an extensive collection of complex processes. The balance between model complexity and resolution, subject to computational limitations, represents a fundamental query in the development of a LSM. With the purpose of adapting the value of the model parameters to values that reproduces results in the real world, measurements are necessary in order to compare to our estimations to the real world. The calibration process consists in an optimization of model parameters for a better agreement between model results and a set of observations, reducing the gap between the model and the available measurements. Here we show how variational data assimilation is applied to the energy and water budgets modules of the ORCHIDEE land surface model in order to constrain the model internal parameters. This part of the model is denoted SECHIBA. The adjoint semi-generator software denoted YAO is used as a framework to implement the 4DVAR assimilation.

A sensitivity analysis was performed in order to identify the most influent parameters to temperature. With the parameter hierarchy resolved, twin experiments using synthetic observations were implemented for controlling the most sensitive parameters. Results obtained suggest that land surface temperature assimilation has the potential of improving the output estimations by adjusting properly the control parameters. Finally, several assimilations were made using observational meteorology dataset from the SMOSREX site in Toulouse, France. The experiments implemented, using different prior values for the parameters, show the limits of the temperature assimilation to constrain control parameters. Even though variable estimation is slightly improved, this is due to final parameter values are at the edge of a variation interval in the cost function. Effectively reaching a minimum would require allowing the parameters to visit unrealistic values. SECHIBA does not correctly simulates simultaneously temperature and fluxes and the relationship between the two is not always consistent according to the regime (or parameter values that are used). We must therefore work on the physical aspects to better simulate the temperature. Likewise, the parameter sensitivity to temperature is not always sufficient, giving as a result a flat cost function.

Our results show that the assimilation system implemented is robust, since performances results in twin experiments are satisfactory. The coupling between the hydrology and the thermodynamics in SECHIBA must be reviewed in order to improve variable estimation. An exhaustive study of the prior errors in the measurements must be conducted in order to retrieve more adapted weighing terms in the cost function. Finally, the assimilation of other variables such as soil moisture should be performed to evaluate the impacts in constraining control parameters

Remerciements

A deep gratitude to all those who in one way or another were involved in the development process of my work and in the writing of this manuscript. To all the people that during these recent years helped me to overcome the difficulties of this project.

Index

CHAPTER 1 INTRODUCTION

LAND SURFACE MODELS. OBJECTIVES AND ORGANIZATION OF THE THESIS	13
1.1 INTRODUCTION	13
1.2 COMPONENTS OF LAND SURFACE MODELS	14
1.2.1 WATER PROCESSES	14
1.2.2 SOIL THERMODYNAMICS	17
1.3 IMPORTANCE OF REPRESENTING THE PHYSICS OF THE SOIL SURFACE CORRECTLY	23
1.4 CHALLENGES IN LSM REPRESENTATION	23
1.4.1. SURFACE HETEROGENEITY	23
1.4.2 NUMERICAL REPRESENTATION	24
1.4.3. MATHEMATICAL REPRESENTATION AND MODEL CALIBRATION	24
1.5 THESIS CHALLENGES	26
1.5.1. STATE OF THE ART IN THE USE OF LST TO CONSTRAIN LSM	26
1.5.2 GENERAL OBJECTIVES	27
1.5.3 ORGANIZATION	28

CHAPTER 2

DESCRIPTION OF THE LAND SURFACE MODEL ORCHIDEE AND DATASETS	31
2.1. ORCHIDEE	31
2.1.1 MODULES	31
2.1.2 BIOSPHERE CHARACTERIZATION	32
2.2 SECHIBA	34
2.2.1 FORCING	34
2.2.2 ENERGY BUDGET	35
2.2.3 HYDROLOGICAL BUDGET	37
2.2.4. SECHIBA PARAMETERS	38
2.3 DATA	40
2.3.1. EDDY COVARIANCE MEASUREMENTS	40
2.3.2 SMOSREX	44

CHAPTER 3

THEORETICAL PRINCIPLES OF VARIATIONAL DATA ASSIMILATION	47
3.1 INTRODUCTION AND NOTATION	47
3.2 ADJOINT METHOD	50
3.3. REPRESENTING A MODEL AND ITS ADJOINT THROUGH A MODULAR GRAPH	51
3.3.1. DEPLOYMENT OF A MODULAR GRAPH	54
3.4. DIAGNOSTIC TOOLS FOR THE ASSIMILATION SYSTEM	55
3.4.1 TEST THE CORRECTNESS OF THE ADJOINT MODEL	55
3.4.2. TEST THE CORRECTNESS OF THE COST FUNCTION GRADIENTS	56
3.4.3. DERIVATIVE TEST	56
3.5. SUMMARY	57

CHAPTER 4

THE YAO APPROACH: THEORETICAL ASPECTS AND IMPLEMENTATION OF SECHIBA-YAO 1D	59
4.1 INTRODUCTION	59
4.2 YAO APPROACH.....	60
4.3. CREATING A PROJECT WITH YAO	61
4.3.1 INPUT / OUTPUT MANAGEMENT	62
4.3.2 DIAGNOSTIC TOOLS FOR THE GENERATED PROJECT	63
4.4. DEVELOPMENT OF SECHIBA-YAO 1D	63
4.4.1 IMPLEMENTATION OUTLINE.....	65
4.4.2 DIRECT MODEL VALIDATION	68
4.4.3 ADJOINT MODEL VALIDATION.....	72
CHAPTER 5	
SENSITIVITY ANALYSIS OF THE SECHIBA-YAO 1D MODEL USING FLUXNET DATASET	77
5.1 INTRODUCTION	77
5.2 VARIATIONAL SENSITIVITY ANALYSIS	79
5.2.1. SENSITIVITY ANALYSIS WITH LAND SURFACE TEMPERATURE.....	79
CHAPTER 6	
TWIN EXPERIMENTS WITH SECHIBA-YAO 1D USING FLUXNET MEASUREMENTS.....	101
6.1 INTRODUCTION	101
6.2 EXPERIMENT DEFINITION	101
6.3. RESULTS	103
6.3.1 EFFECT OF THE OBSERVATION SAMPLING	103
6.3.2 EFFECT OF RANDOM NOISE IN THE OBSERVATION	105
6.3.3 EFFECT OF THE CONTROL PARAMETER SET SIZE.....	106
6.4. DISCUSSION.....	109
CHAPTER 7	
REAL MEASUREMENTS STUDY USING SMOSREX DATASET	111
7.1 INTRODUCTION	111
7.2 KEY PARAMETERS TO PERFORM THE OPTIMIZATION	112
7.3 LST DATA ASSIMILATION WITH PARAMETER STANDARD VALUES	113
7.3.1. SIMULATED VS. OBSERVED MEASUREMENTS	113
7.3.2. BRIGHTNESS TEMPERATURE SENSITIVITY ANALYSIS.....	119
7.3.3. BRIGHTNESS TEMPERATURE ASSIMILATION DURING A SINGLE DAY.....	120
7.3.4. BRIGHTNESS TEMPERATURE ASSIMILATION DURING A WEEK	122
7.3.5. DISCUSSION	126
7.4 LST VARIATIONAL DATA ASSIMILATION WITH DIFFERENT PRIOR VALUES	126
7.4.1. SIMULATED VS. OBSERVED MEASUREMENTS	126
7.4.2. BRIGHTNESS TEMPERATURE SENSITIVITY ANALYSIS.....	129
7.4.3. BRIGHTNESS TEMPERATURE ASSIMILATION DURING A SINGLE DAY.....	130
7.4.4. BRIGHTNESS TEMPERATURE ASSIMILATION DURING A WEEK	132
7.5. ANALYSIS OF THE ASSIMILATION SYSTEM THROUGH TWIN EXPERIMENTS	133
7.6 CONCLUSION	

CHAPTER 8

CONCLUSION AND PERSPECTIVES	139
REFERENCES	145
FIGURE INDEX	151
TABLE INDEX	155
APPENDIX A	157
APPENDIX B.....	161
APPENDIX C.....	171

Chapter 1 Introduction

Land Surface Models. Objectives and Organization of the Thesis

1.1 Introduction

A land surface model (LSM) is a numerical model describing the exchange of water and energy between the land surface and the atmosphere. It is used in weather and climate modeling to simulate different processes at the Earth's surface, such as the water, carbon fluxes and the thermodynamics from the surface to the atmosphere. The atmospheric model provides the forcing above the surface (incoming radiation, precipitation, atmospheric temperature and humidity). In exchange, LSM calculates surface variables (soil temperature, soil moisture, leaf area index, etc...) and outgoing fluxes to the atmosphere.

A good description of surface processes is essential for weather and climate applications. Located at the boundary between the atmosphere and the soil, LSM provides the link between several scientific disciplines subject to intense research in the hydrological, atmospheric, and remote sensing communities. LSM contains a set of general components, as mentioned in Liu and Gupta, 2007. First, we have the system boundary, which sets the limit conditions to model variables, allowing the separation of internal components of a system from external entities. Second, forcing inputs describe the model variables modified by the external components of the model: these are the inputs of the model. Next, the initial states of the model define a more realistic scenario for the state variables. We also have the model parameters, which are part of the equations describing the physical phenomena. Finally, the model structure contains a description of how the model is decomposed in its elementary processes and the discretization path used.

Principal components of LSM and their associated physics are explained in the next section.

1.2 Components of land surface models

1.2.1 Water processes

The soil surface regulates the exchanges between water and plants. The ability of soil to hold water is related to its constitution. It is composed of particles that can be large or compact. Depending on soil composition, its ability to absorb water is different. Very dry soil cracks and becomes very compact. It cannot store water, which escapes through the large cracks on the surface, and runs off. Therefore in such areas the vegetation is limited or nonexistent. The water balance defines the soil water content by integrating all the water that comes in and all the water that leaves the soil (Musy et al., 1991).

All movements of water in the soil depend on its structure and its state: the infiltration process consists in vertical water movement that enters the soil through tiny cracks together. This phenomenon occurs in the first meters below the surface of the Earth. This movement maintains the reserves of deep water. The infiltration process modifies in a drastic and instantaneous way the pressure and the water content in the ground surface.

This process is conditioned by several factors of which the most significant comes from the soil, through its hydrodynamics characteristics, its texture and its structure. In addition, specific conditions can determine the infiltration process: the water flow rate, precipitation intensity, etc.

Hydrological Cycle

The hydrological cycle is a concept that encompasses the phenomena of movement and renewal of water on Earth. Concepts in this section are taken from Musy et al., 2003. This definition implies that the mechanisms governing the hydrological cycle does not only occur one after the other, but there are also concurrent. Under the effect of solar radiation, water evaporates from the soil, oceans and other surfaces such as lakes, rivers, etc. The rise of moist air through the atmosphere facilitates its cooling, which is necessary to bring it to saturation, causing condensation of water vapor into droplets forming clouds in the presence of condensation nuclei. Then the water vapor, transported and stored temporarily in the clouds, is returned to oceans and continents through rainfall. Part of the rainfall can be intercepted by plants and then be partially restored by evaporation. Intercepted rain not evaporated can

further reach the ground (NOAA National Weather Service Jetstream).

Depending on soil conditions, water might evaporate from the ground, run off, infiltrate into the ground where it can be stored as soil moisture or leak into deeper areas to contribute to the renewal of groundwater reserves. Flow from the latter can reach the surface at springs or streams. Evaporation from the soil, rivers, and plant transpiration complete the cycle.

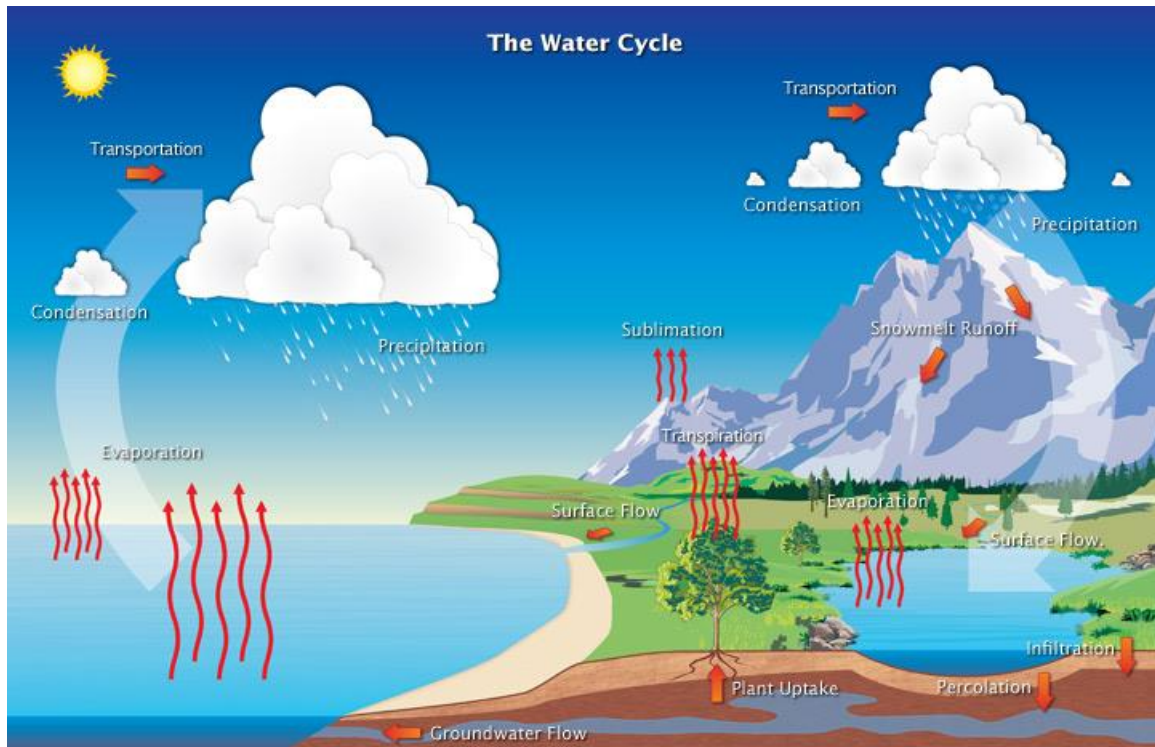


Figure 1. 1 Water moves through the Earth, changing state and drifting to the atmosphere, the oceans and over the land surface and underground, in different processes that coexist. It is subject to complex processes; among them we cite precipitation, evaporation, transpiration, interception, runoff, infiltration, percolation, storage and subsurface flows. These various mechanisms are made possible by the incoming surface energy. (Source: NOAA National Weather Service Jetstream).

Water Balance

We can represent the continuous process of the water cycle in three general phases: precipitation, surface runoff, infiltration and evaporation. In each phase we found a transport of water and sometimes a change of state. The water balance equation takes these phases into account, and it is expressed as follows

$$S = P - R - E \quad [\text{mm}] \quad (1.1)$$

With P the precipitation (liquid and solid), S the resources (accumulation) in the previous period (infiltration, soil moisture, snow, ice), R the surface runoff and groundwater

flow and E the evapotranspiration, defined as the integration of evaporation and transpiration.

Precipitation is the rain water falling on the surface of the earth, both in liquid and solid form. In order to produce condensation in the atmosphere, water vapor must reach the dew point by cooling or by increasing its pressure. In addition, the presence of certain microscopic nuclei allows water droplets to condense. The source of these nuclei may be oceanic (chlorides, in particular $NaCl$ produced by sea evaporation), continental (dust, smoke and other particles carried out by ascending air currents) or cosmic (meteoric dust). The onset of precipitation is favored by the coalescence of water drops. The increase in mass gives them a gravitational force sufficient to overcome the updrafts and turbulence of the air and to reach the ground. Finally, the course of raindrops or snowflakes must be short enough to prevent evaporation of the total mass.

Evaporation is defined as the transition from liquid to vapor. Water bodies, such as lakes and oceans and vegetation cover are the main sources of water vapor. The main factors governing the evaporation are the solar radiation, the wind and the soil moisture. The term evapotranspiration includes evaporation and plant transpiration. It is a fundamental component of the hydrological cycle and its study is essential to determine the water resources of a region or watershed. In general, specific analysis of evaporation will be made to balance studies and management of water by plants. Evaporation of the soils is produced at the surface. Transpiration is produced by plants essentially by the leaves with water extracted from the soil in the root zone. Both processes occur simultaneously. The evapotranspiration process is conditioned by the evaporative power, which expresses the extraction capacity of water, performed by the atmosphere on the ground-vegetation system. This evaporative power is determined by the evaporative demand of the air and the system's ability to satisfy this request, depending on the availability of water and plant physiology.

Infiltration refers to the downward movement of water from the land surface into the soil. Infiltration can occur naturally following precipitation, or can be induced artificially through structural modifications in the ground surface (Musy et al., 1991). The infiltration process is characterized by the water influx in the soil. This influx is called infiltration regime, which is defined also as the flux density q . The infiltration rate can be influenced by the soil properties like its porosity, the hydraulic conductivity and the initial humidity rate.

The precipitation penetrating the soil creates several zones, as shown in Fig.1.2. These zones are: the saturation zone at the soil surface; the transmission zone, which is characterized by constant water content; the wetting zone, in which the water content is decreased and ends

up in the last zone and finally the wetting front, characterized by an abrupt decrease in the water content.

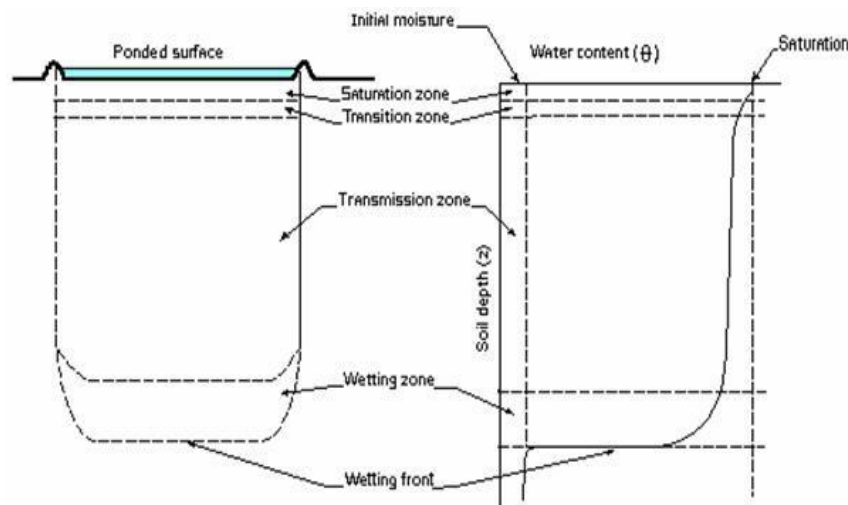


Figure 1. 2 Soil profile and the different types of saturation zones. (Musy and Soutter, 1991).

The soil water capacity is expressed as a depth of water that can be infiltrated per unit of time. If rainfall supplies water at a rate that is greater than the infiltration capacity, water will infiltrate at the capacity rate, with the excess either being ponded, moved as surface runoff, or evaporated. If rainfall supplies water at a rate less than the infiltration capacity, all of the incoming water volume will infiltrate. In both cases, as water infiltrates into the soil, the capacity to infiltrate more water decreases and approaches a minimum capacity. When the supply rate is equal to or greater than the capacity to infiltrate, the minimum capacity will be approached more quickly than when the supply rate is much less than the infiltration capacity.

1.2.2 Soil thermodynamics

Formalism and notations presented in Musy and Soutter (1991) are adopted in this part of the work. Soil thermodynamics describes the different processes representing soil temperature in space and time. Temperature is a variable characterizing the degree of thermal distribution and the level of body heat. Many physical, chemical and biological processes are strongly influenced by temperature. The occurrence of thermal gradients results in transfers by heat diffusion, liquid flows, soil aeration, drainage and exchanges with the atmosphere. They are all conditioned by the thermal state of the soil. The soil temperature at a specific point depends on two types of phenomena: on one hand the energy exchange with the outside environment, determining the amount of total energy stored in the ground, and on the other, a

complex series of transfer processes of heat, depending on the specific thermal properties of soil, which determines energy distribution in the ground.

Energy transfer

Heat transfers are energy exchanges, following three types of energy physical processes: radiation, thermal and latent. They are defined based on the formulation of their dynamics and a conservation law. The main parameters involved in the description of the thermal behavior of a soil are those used to characterize the stored energy. These values depend on the specific thermal properties of the various components of the soil.

The thermal properties of the soil are conditioned by the characteristics of their texture and structure. One of these parameters is C_t , the heat capacity of a body, which is defined by $\frac{dQ}{dT}$, where dQ is the energy-heat required to raise the body temperature per dT . The isobaric heat capacity is denoted by

$$C_p = \frac{C_r}{M} \quad (1.2)$$

where C_p is the storage capacity of body heat per unit mass and temperature. Internal exchanges of energy, thermal energy or sensible heat transfer in soil occur in two different processes: thermal conduction and convection.

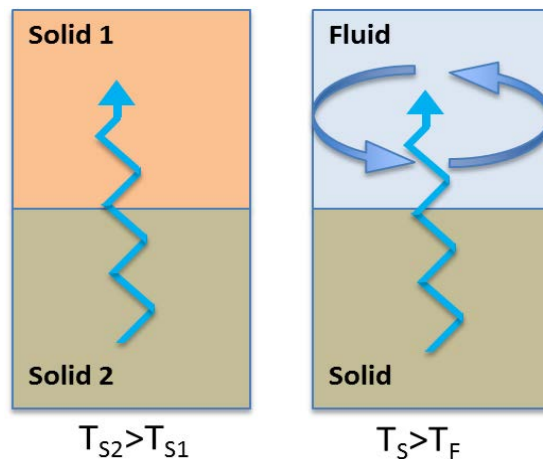


Figure 1. 3 Thermal conduction between two solid (left) and thermal convection between a solid and a fluid (right). (Source: Musy and Soutter, 1991)

The thermal energy exchange in soil generally occurs simultaneously by thermal

conduction and convection, denoted as \vec{J}_D and \vec{J}_v , respectively (Fig.1.3). Both are based on different physical exchange heat processes. Their dynamics is described by different laws, depending on the number of phases involved. The total heat flux is defined as

$$\vec{J}_T = \vec{J}_D + \vec{J}_v \quad (1.3)$$

Thermal conduction is a diffusion process in which the transfer of energy is due to the difference in temperature between two regions of a medium or between two media in contact without displacing material. The thermal agitation is transmitted from one molecule or atom giving a portion of its kinetic energy to its neighbor. This transfer process ends only when a heat balance is reached. The heat flux transferred by conduction is proportional to a gradient of decreasing temperature, defined by the Fourier law:

$$\vec{J}_D = -K_T \text{grad } T \quad (1.4)$$

The Fourier coefficient or thermal conductivity K_T in Eq.1.4 represents the resistance of the material to the propagation of heat by thermal conduction, expressing its ability to transfer heat from one point to another. Thermal conductivity depends on the composition of the material, its mineral content and organic matter as well as the shape of its constituent particles. It varies in space and in time as a function of variations in air moisture.

In thermal convection, the transfer of energy takes place in the molecules located at the boundary between a solid and a fluid in motion. The transfer is carried out by the molecular agitation and fluid mass displacement. This heat transfer is expressed in terms of δT_i , the thermal density, is described as

$$\delta T_i = \rho_w c_{pw} (T_S - T_W) \quad (1.5)$$

Where ρ_w is the water density and c_{pw} is the isobaric heat capacity density. The exchanged heat by the movement of the fluid must be associated with the law expressing the fluid dynamics. The sensible heat flux transferred by thermal convection \vec{J}_v is written using the following expression

$$\vec{J}_v = \delta T_i \vec{q} \quad (1.6)$$

In natural convection, heat is transferred by density currents that run through the fluid under the effect of disparities in temperature at the solid-fluid interface. In this case, the

sensible heat flux is expressed as

$$\vec{J}_v = -\rho_w c_{pw} D_T \text{grad } T \quad (1.7)$$

where D_T is the molecular diffusivity. The temperature gradient determines the transfer of thermal energy from one phase to the other. The dynamic law expressing the total sensible heat flux results from the sum of conduction and convection transferred by heat flow.

The general equation of exchange of sensible heat is the combination of the dynamic law and the principle of continuity. It can be expressed in terms of thermodynamics principles.

$$\frac{\partial \theta}{\partial t} = \text{div}(D \text{grad } \theta) + \text{div}(D_x \text{grad } T) + \frac{\partial K}{\partial z} \quad (1.8)$$

$$\rho_w c_{pw} \frac{\partial T}{\partial t} = \text{div}(K_T \text{grad } T) - \lambda \text{div}(D_\lambda \text{grad } \theta) \quad (1.9)$$

The first equation (Eq.1.8), includes the transfer of water under the influence of the potential gradient of water and the thermal gradient. Here, D is the water transfer diffusivity, D_x is the thermal gradient diffusivity and θ is the quantity of water in the soil. The second equation (Eq.1.9) is the general equation of heat flux to which we add a latent heat flux due to evaporation flow. Here, D_λ and λ are terms related to the latent heat flux diffusivity of soil.

Energy budget

The sun is the main source of energy reaching the surface of the Earth. In a single year, the Earth system (atmosphere, surfaces, and oceans) absorbs sunlight driving photosynthesis, evaporation, and melts snow and ice, among other processes. The heat collected is not uniform across surface of the Earth. The atmosphere and ocean balance the energy received by the Sun through surface evaporation, rainfall, ocean circulation, winds, etc. Mean temperature in Earth doesn't increase relentlessly because the surface and the atmosphere are simultaneously radiating heat back to space. This net flow of energy is known as the energy budget.

Variations in the average temperature reflect the balance of energy exchanges between the soil and the outside environment. These exchanges occur at the soil-atmosphere interface in the form of radiant energy, heat and latent.

The net radiation is the difference between the amount of incoming radiation and the amount of outgoing radiation by the surface. Reflectivity of the atmosphere and the ground

surface determines the amount of radiation the surface will absorb. This reflectivity is known as albedo, a coefficient that determines the reflected solar radiation for a particular surface. Albedo varies with surfaces, leading to net heating inequities throughout the Earth: more incoming sunlight is received in the summer hemisphere compared to the winter hemisphere. When the Earth is subject to a flow of radiant energy, it absorbs a portion, reflects another and transmits the remainder, as it is described in Figure 1.4.

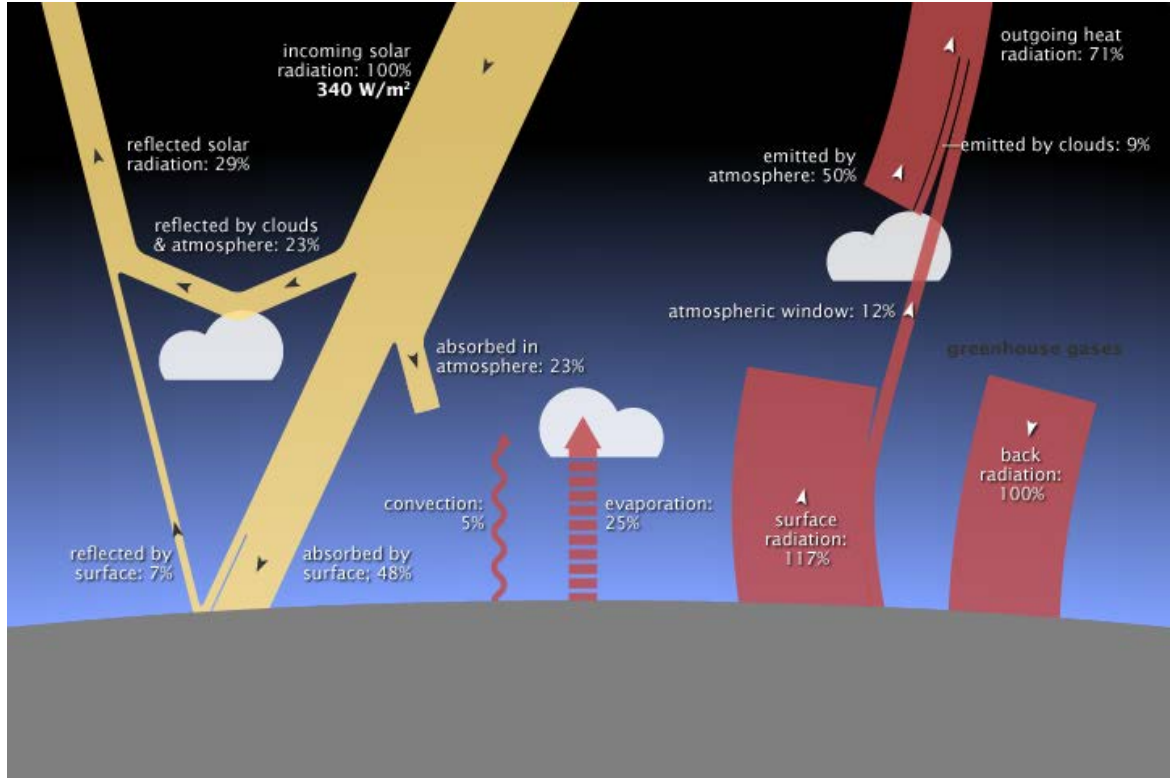


Figure 1. 4 Global Earth Energy Budget. (Source: NASA)

The energy radiation budget exchange to the surface is expressed by the net radiation, which is represented by

$$R_n = (R_h \downarrow - R_h \uparrow) + (R_b \downarrow - R_b \uparrow) \quad (1.10)$$

Where $R_h \downarrow$ is the incident solar radiation, $R_h \uparrow$ is the reflected solar radiation, R_b is the radiation emitted by the surface. R_n can be expressed also as the global radiation R_g product with the surface albedo α , following the equation

$$R_n = SW \downarrow (1 - \alpha) - \varepsilon \sigma T^4 + \varepsilon LW \downarrow \quad (1.11)$$

Another element of the energy budget is represented by the exchange of sensible heat. The exchange of thermal agitation energy by convection is the main process of sensible heat

transfer between the ground and the atmosphere. The convection exchange takes place simultaneously by natural convection (heat distribution) and by forced convection, as a result of the mechanical stirring action of the wind. The sensible heat flux H is represented by

$$H = -\rho_a C_{pa} D_T \text{grad} T \quad (1.12)$$

where ρ_a is the air density, C_{pa} is the isobaric specific heat capacity and D_t is the thermal diffusivity. Another type of exchange occurs in the form of latent heat, wherein the energy has been converted by a phase change of heat transfer and thus being associated with a mass transfer. Sensible heat converts during this process to latent heat, transferred by the vapor mass flow. This flow can be expressed in general as the product of the concentration gradient, the volume fraction coefficient V_{vap} and an exchange coefficient D_e . Diffusive and convective effects transfer of water vapor in the air are then defined as

$$\vec{q}_{vap} = -D_e \text{grad} V_{vap} \quad (1.13)$$

The latent heat flux associated to the water vapor mass flow is expressed as

$$LE = -\frac{\rho_a C_{pa}}{\gamma} D_e \text{grad} p_{vap} \quad (1.14)$$

where LE is the vaporization latent heat, p_{vap} is the evaporation pressure and γ is the psychrometric constant, defined as $\gamma \approx 0.66.10^{-2} \text{ N/m}^2\text{K}$

Formulation

The energy budget provides the balance between the balance of radiation and heat exchange. It is calculated from the following equation:

$$R_n = M + H + LE + G \quad [\text{W/m}^2] \quad (1.15)$$

R_n is the net radiation, M represents the part of the radiation energy absorbed by the system and used for photosynthesis, and it is usually neglected. G represents the radiation energy converted into heat and stored in the system, after deducting the sensible heat flux H and latent heat flux LE .

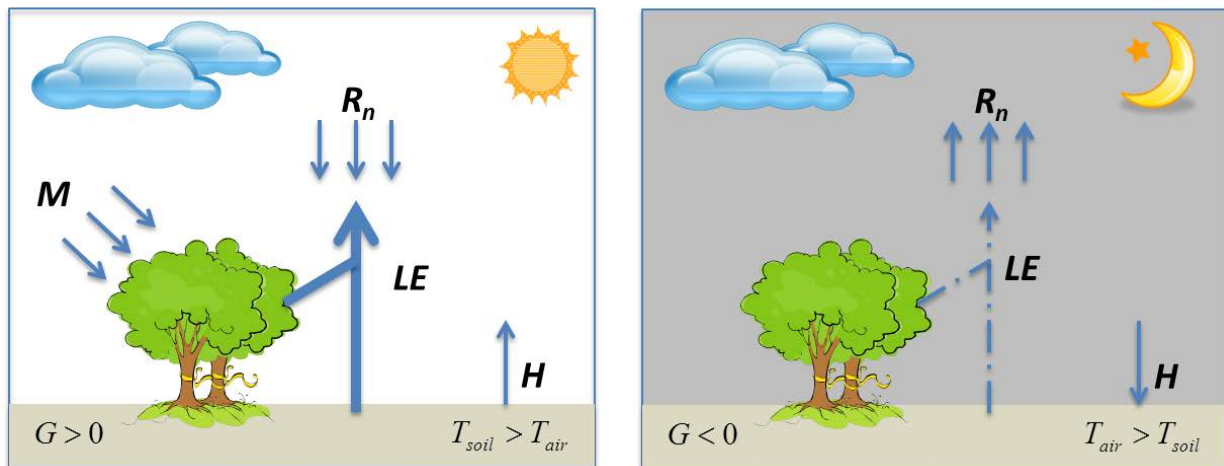


Figure 1. 5 Daylight and night energy budgets. (Source: Musy et Soutter (1991)).

1.3 Importance of representing the physics of the soil surface correctly

Land surface model helps us to understand the processes that trigger the precise climate in a region, by analyzing its internal interactions. We are able to understand the current climate locally but also the factors that generate a specific climate in other regions. Fluxes between soil and the atmosphere can have non-local effects. In addition, climatic conditions into the future can be predicted. Its accuracy depends on the degree of precision with which the model represents the physics reality and the assumptions about the future factors that climate will encounter.

Furthermore, LSM provides data not directly available or difficult to interpolate. Models compute different estimates answering questions about what processes will affect a particular region in the future.

1.4 Challenges in LSM representation

The land surface physics includes an extensive collection of complex processes. The balance between model complexity and resolution, subject to computational limitations, represents a fundamental question in the development of LSM. By increasing the comprehension of physical phenomena, LSM can grow into a more complex model adding new processes extrapolated from the environment. In the next section, some of the complexities found in most LSM are presented.

1.4.1. Surface heterogeneity

The soil is a complex environment. Its composition includes inorganic and organic

particles that determine its inherent characteristics. Its ability to retain water and its density varies drastically depending on the percentage of primitive components (sand, silt, or clay) or structure (ash, fine, or coarse sand). There may be several types of soil in a small area, with their associated features, especially water content. Omission of surface spatial heterogeneity in a LSM can cause errors in flux estimation (Courault et al., 2010, Oliso et al., 2005). Spatial variations in surface heterogeneity are imperative in order to guarantee an accurate simulation of the land-surface fluxes. The sub-grid scale land surface heterogeneity must be parameterized in the surface scheme so that the land characteristics are accounted for in the model (Manrique et al, 2013).

1.4.2 Numerical representation

The development process of a physical model into numerical software extends from the physical world to the mathematical model, then to the computational algorithm and finally to the computer implementation, involves a number of approximations: physical effects may be discarded, continuous functions replaced by discretized ones and real numbers replaced by finite precision representations. In consequence, approximation is in the core of scientific software and cannot be neglected. It is important to manage them judiciously.

The accuracy of a computation determines how close the computation (affected by random errors) comes to the true mathematical value. It indicates, therefore, the correctness of the result. In particular, numerical verification, as Rump (1983) mentioned, is required to give confidence that the computed results are acceptable. The precision of a computation reflects the exactness of the computed result without reference to the meaning of the computation. It is, therefore, the number of significant digits affected by round-off error. Arithmetic expressions and variable assignment always produce approximation errors, due to the nature of the floating point arithmetic. Approximation modes in computer software will determine the precision in the several operations made in the model coded. This precision is independent of the code, data or machine. When building LSM numerical representation, we have to be aware of these errors and track their propagation.

1.4.3. Mathematical representation and model calibration

Representing a physical model need the definition of its numerical and its discretized form. In terms of simplicity, models have to include the least amount of parameters needed to achieve a good performance in its estimates (parsimonious models). Model building is best

achieved by starting with the simplest structure and gradually and accurately increasing the complexity as needed to improve model performance (Wainwright et al., 2004). However, there is no metric that quantifies the estimate improvement by increasing complexity.

The parameters that are required to compute real world outputs estimations are best defined by a model structure that best represent the processes measured in the real world. In practice, this can be difficult to achieve. With our model, we may be interested in trying to reconstruct past events that need some parameters that are impossible to measure. In that case we may have to make reasonable assumptions based on indirect evidence. These assumptions can be made in order to define the model from reality. Several of them will and can be wrong, nonetheless they are necessary for the model development. The output of the model depends completely on the validity and scope of these assumptions. A parameter measurement must be chosen based on the impact the variation of a parameter has on the model output, or the model sensitivity to this parameter.

According to Kirkby et al (1992) there are two types of parameters in a model: the physical parameters which define the physical structure of the model, and the process parameters or multiplying factors, which weigh the magnitude of variables in the model. The physical parameters are determined from experimental measurements. The process parameters are defined from a calibration and adjustment process. In both cases, the definition of the initial parameter value can be a difficult task. The physical parameters are determined on small scales, and then they are extrapolated, given the spatial and temporal variability in the region we are working.

With the purpose of adapting the value of the parameter to a value that reproduces the real world, measurements of a phenomenon are necessary in order to compare to our model estimates. The calibration process consists in an optimization process against a measure of the agreement between model results and a set of observations. It allows the agreement between the model and the available measures; however, this process may give clues to poorly defined processes in the model (Pipunic et al., 2008).

With respect to the LSM, many works have focused on the calibration of the models based on soil moisture measurements, since it is an observation easy to obtain, it is directly measured with high frequency and is the solution of the water budget. There are many sources of data available in a wide range of ecosystem.

1.5 Thesis Challenges

1.5.1. State of the art in the use of LST to constrain LSM

Several works regarding the calibration of LSM based on LST measurements demonstrate the improvement in fluxes estimation, when constraining model parameters.

In Castelli et al. (1999), a variational data assimilation approach is used to include surface energy balance in the estimation procedure as a physical constraint (the adjoint technique). The authors work with satellite data, where soil skin temperature is directly assimilated. As a conclusion, constraining the model with such observation improves model fluxes estimations, with respect to *in situ* measurements.

In Huang et al. (2003) the authors developed a one-dimensional land data assimilation scheme based on ensemble Kalman filter, used to improve the estimation of soil temperature profile. They conclude that the assimilation of LST into land surface models is a practical and effective way to improve the estimation of land surface state variables and fluxes.

Reichle et al. (2010) performed an assimilation of satellite-derived skin temperature observations using an ensemble-based, offline land data assimilation system. Results suggest that retrieved fluxes provide modest but statistically significant improvements. However, they noted strong biases between LST estimates from *in situ* observations, land modeling, and satellite retrievals that vary with season and time of day. They highlighted the importance to take these biases properly, or else large errors in surface flux estimates can result. In Ghent et al. (2011), the authors investigate the impacts of data assimilation on terrestrial feedbacks of the climate system. Assimilation of LST helped to constrain simulations of soil moisture and surface heat fluxes. Another study by Ghent et al. 2011, investigates the effect that data assimilation has on terrestrial feedbacks to the climate system. The authors state that representation of highly complex biophysical processes in LSMs over highly heterogeneous land surfaces with limited collections of mathematical equations, and the tendency of over parameterization, infers a degree of uncertainty in their predictions. Assimilation of land surface temperature (LST) to constrain simulations of soil moisture and surface heat fluxes can be integrated into the model to update a quantity simulated by the model with the purpose of reducing the error in the model formulation. The correction applied is derived from the respective weights of the uncertainties of both the model predictions and the observations. The results found in this research suggest that there is potential for LST to act as surrogate for assimilating other state variables into a land surface scheme.

Ridler et al. (2012) tested the effectiveness of using satellite estimates of radiometric surface temperatures and surface soil moisture to calibrate a Soil–Vegetation–Atmosphere Transfer (SVAT) model, based on error minimization of temperature and soil moisture model outputs. Flux simulations were improved when the model is calibrated against *in situ* surface temperature and surface soil moisture versus satellite estimates of the same fluxes.

In Bateni et al. (2013), the full heat diffusion equation is employed as constrain, in the variational data assimilation scheme. Deviations terms of the evaporation fraction and a scale coefficient are added as penalization terms in the cost function. Weak constraint is applied to data assimilation with model uncertainty, accounting in this way for model error. The cost function in this experience contains a term that penalizes deviation from prior values. When assimilating LST into the model, the authors proved that the heat diffusion coefficients are strongly sensitive to specific deep soil temperature. As a conclusion, it can be seen that the assimilation of LST can get a remarkable improvement in the model simulated flows.

1.5.2 General objectives

In this work, the LSM used is ORCHIDEE (Krinner et al., 2005), most specifically the part of the model computing the energy and hydrology balance (SECHIBA, Ducoudré et al, 1993). These models are introduced in Chapter 2.

The general objective of this thesis is to constrain the SECHIBA model parameters by assimilating measurements products in a 4DVAR assimilation system. The parameters, once constrained, allow the model to improve state variables estimation when comparing them to measurements.

From this general purpose, several specific objectives arise, as mandatory steps to implement an effective assimilation system, flexible enough to assimilate different observations and constraining at the same time different model parameters. These specific objectives are:

1. Study of SECHIBA and implementation into YAO: the understanding of the model physics through its standard Fortran code implementation is a mandatory step, in order to extract model dynamics and principal components. By knowing this, the implementation of SECHIBA in YAO can be made, by defining a modular graph representing the model dynamics and physics of the model. Our implementation of SECHIBA in YAO is called SECHIBA-YAO 1D. Once our model is coded, the direct model is verified comparing its

output with the original model. The adjoint model is verified by performing a sensitivity analysis, allowing us to obtain, in addition, a parameter hierarchy of the most influential parameter in the estimation of land surface temperature. SECHIBA-YAO 1D aims to run 4DVAR assimilation.

2. Validate the assimilation system, by implementing twin experiments. The idea is to test the robustness of the assimilation system, by computing variable and parameter performances. This phase highlights also the limits of the model when varying the control parameter set
3. Improve model estimation by performing a 4D-VAR assimilation of land surface temperature, using *in situ* measurements of SMOSREX site, in Toulouse, France. Available measurements of brightness temperature are compared with an equivalent form of temperature estimation added to SECHIBA, constraining model parameters to improve the simulation of the model variables, such as latent heat flux, sensible heat flux, net radiation, brightness temperature and soil moisture.

1.5.3 Organization

The thesis is organized such that the theoretical support is presented first, in Chapters 1 to 4. Experiment results concerning sensitivity analysis and variational data assimilation are presented in Chapters 5 to 7. Conclusions of the thesis are presented in Chapter 8. Finally, complementary information is presented in the Appendix section.

In Chapter 1, the introduction to land surface models and the nub of the thesis is presented. In Chapter 2, the land surface model used in this work (ORCHIDEE) is introduced, and more specifically SECHIBA and its main components and features. Equations governing the energy and hydrologic budget computed with SECHIBA are listed. In addition, data sources used in this work are introduced: FLUXNET network stations and SMOSREX *in situ* measurements.

Chapter 3 concerns variational data assimilation theoretical aspects. Additionally, the modular graph approach to represent models is presented. It is explained how an equivalent of the adjoint and tangent linear model is obtained, by computing the forward and the backward of the model through a modular graph decomposition. This approach is the basic idea of the YAO software, serving as a framework to implement SECHIBA variational data assimilation.

In Chapter 4, the YAO approach is presented. This software served us as an adjoint semi

generator. Principal components of a YAO project are introduced, as well as the input/output data management. Finally, a general guide of how SECHIBA model was implemented in YAO is introduced. The different steps from the model conceptualization to the testing phase are explained in detail, serving as a guide to future implementations.

In Chapter 5, once the adjoint of SECHIBA is obtained using its YAO representation, we perform a variational sensitivity analysis with the idea of validating the adjoint of the model, by comparing the gradients obtained with SECHIBA-YAO 1D to the ones computed with finite differences, using the direct model outputs. In addition, we build a parameter hierarchy in order to determine the most influential parameters in the computation of land surface temperature. The sensitivity study was performed using FLUXNET sites (Kruger Park and Harvard Forest).

In Chapter 6, twin experiments are presented, using the FLUXNET data set. Different scenarios were tested in different experiments in order to account for the effect of assimilating synthetic observations of land surface temperature. In this chapter we show the potential of our assimilation system by using land surface temperature as observation.

In Chapter 7, assimilation of *in situ* measurements is presented using the SMOSREX site forcing. With different scenarios, this chapter shows the performance of assimilating land surface temperature with different initial conditions of parameters, time frames, among others. The idea is to show if the optimization of land surface temperature allows to constrain our model parameters and to better simulate surface fluxes.

Finally, in Chapter 8, the conclusion and perspectives of this work are presented. In addition, challenges related to the implementation of SECHIBA into YAO are mentioned.

Chapter 2

Description of the Land Surface Model ORCHIDEE and datasets

2.1. ORCHIDEE

ORCHIDEE is a model representing the continental biosphere and its different processes, comprising the simulation of soil and vegetation mechanism and simulating different fluxes between the soil-atmosphere interface (Polcher et al., 1998, Krinner et al., 2005, Brender, 2012). ORCHIDEE has different time scales: energy and matter has a 30-minutes time scale. Species competition processes at 1-year time scale. The vegetation is grouped into 13 Plant Functional Type (PFT). The equations governing the processes are general, with specific parameters for each PFT. ORCHIDEE is used in a grid-point mode (one given location), forced with the corresponding local half hourly gap-filled meteorological measurements.

2.1.1 Modules

SECHIBA (Schématisation des Echanges Hydriques à l'Interface Biosphère-Atmosphère) (Ducoudré et al, 1993) is a biophysical model. It calculates the radiation and energy budgets of the surface, and the soil water budget every half hour. The energy and water fluxes between the atmosphere and the ground integrate all the vegetation layers; the retrieved temperature represents the canopy ensemble and the soil surface. The main fluxes modeled are the sensible and latent heat flux between the atmosphere and biosphere, the soil temperature and the water reservoirs evolution, the stomata conductance and gross primary productivity of the canopy.

STOMATE (Saclay Toulouse Orsay Model for the Analysis of Terrestrial Ecosystems) is a biogeochemical model. It represents the process related to the carbon cycle, such as carbon dynamics, the allocation of photosynthesis (Friedlingstein et al, 1999), respiration and growth maintenance, heterotrophic respiration (Ruimy et al., 1993) and phenology (Botta, 1999). STOMATE simulates the dynamics of continental carbon with no time every day. It links between processes at short time scales determined by SECHIBA and slower processes described by the following module.

LPJ (Lund-Potsdam-Jena) (Sitch et al, 2003) is a model of global dynamics of the vegetation. It incorporates the phenomena of interspecific competition for sunlight, fire occurrence, seedling establishment, plant mortality, and deduce the dynamic long-term (annual time step) of vegetation.

2.1.2 Biosphere Characterization

The surface model SECHIBA aims at representing the water and energy exchanges at the land surface. However, for a given moisture condition, they are highly dependent on soil type and vegetation cover. ORCHIDEE considers the diversity of a given ecosystem by defining 13 Plant functional Type (PFT). The vegetation is classified according to their ecophysiologic characteristics. Twelve common PFT exists, plus the bare soil; they are presented in Table 2.1. This classification depends on several parameters such as the appearance of the plant (tree or herb), the type of leaf (needle or leaves), the method of photosynthesis (C3 or C4) and the phenology type.

The different functional groups of plants and bare soil can coexist on the same mesh. A PFT is not intended to represent a plant species in particular but rather to regroup it, after several functional similarities. This will set the main functional characteristics such as height, *LAI*, etc; and thus represent plant diversity around the world.

LAI (Leaf Area Index) is the total ratio of leaf area of a canopy over an area of soil. It is expressed in m² of leaf area per m² of soil. For each mesh point, a fraction of PFT f_k is defined as the vegetation fraction percent covering the studied location, verifying that $\sum_{k=1}^{13} f_k = 1$. Each fraction is modulated by a maximum f_k^{\max} and by a corresponding *LAI* with the following equation:

$$f_k = f_k^{\max} \cdot \min(2 \cdot LAI_k, 1) \quad k = 2, 13 \quad (2.1)$$

$$f_1 = f_1^{\max} + \sum_{k=2}^{13} (f_k^{\max} - f_k) \quad (2.2)$$

If we have a LAI less than 0.5, we reduce linearly the vegetation fraction to zero for a LAI equal to zero and we increase equally the fraction of bare soil. Evolution of a LAI for each PFT is bounded by minimum and maximum values which are assigned in Table 2.1. These values are reached according to the change in soil and vegetation temperatures

PFT	Description	Foliage	Climate
0	Bare Soil	-	-
1	Rain forest <i>sempervirens</i>	Persistent	Tropical
2	Rainforest <i>deciduous</i>	-	Tropical
3	Temperate forest of conifers <i>sempervirens</i>	Persistent	Temperate
4	Temperate forest <i>sempervirens</i>	Persistent	Temperate
5	Temperate forest <i>deciduous</i>	-	Temperate
6	Boreal coniferous forest <i>sempervirens</i>	Persistent	Boreal
7	Boreal deciduous	-	Boreal
8	Boreal coniferous forest <i>deciduous</i>	-	Boreal
9	Herbaceous C3	C3	-
10	Herbaceous C4	C4	-
11	Agricultural C3	C3	-
12	Agricultural C4	C4	-

Table 2. 1 ORCHIDEE's Plant functional type description (d'Orgeval 2006)

In Table 2.2, h represents the prescribed vegetation height in meters, hum_{cste} is the root profile coefficient (in m^{-1}), r_k is the structural resistance (in $s.m^{-1}$), T_{min} and T_{max} are maximum and minimum values of soil temperature at 50 cm and LAI_{max} and LAI_{min} are the maximum and minimum values of LAI for each vegetation fraction.

Energy balance is solved once, with a subdivision only for latent heat flux in bare soil evaporation, interception and transpiration for each type of vegetation. Water balance is computed for each vegetation type given that, for a particular location, infiltration and evaporation will be different.

PFT	h	hum_{cste}	r_k	LAI_{min}	LAI_{max}	T_{min}	T_{max}
0	0	0	0	0	0	275.15	273.15
1	30	0.8	25	8	8	296.15	300.15
2	30	0.8	25	0	8	296.15	300.15
3	20	1	25	4	4	278.15	288.15
4	20	0.8	25	6	6	278.15	288.15
5	20	0.8	25	0	6	278.15	288.15
6	15	1	25	4	4	278.15	288.15
7	15	1	25	0	6	278.15	288.15
8	15	0.8	25	0	4	278.15	288.15
9	0.5	4	2.5	1	5	280.15	288.15
10	0.6	4	2	0	4	284.15	294.15
11	1	4	2	0	6	280.15	288.15
12	1	4	2	0	4.5	284.15	294.15

Table 2. 2 ORCHIDEE's plant functional type principal parameters (d'Orgeval 2006)

2.2 SECHIBA

Our study focuses on the vertical hydrological processes and the energy budget modeled in SECHIBA module. The other two modules of ORCHIDEE (i.e. STOMATE and LPJ) were not active. We chose to make simple assumptions concerning the modeling of vegetation cover. In order to do that, SECHIBA can be used decoupled from STOMATE.

2.2.1 Forcing

SECHIBA uses a time step of 30 minutes to represent the physical processes. The spatial resolution is determined by the atmospheric forcing used. For simulating surface fluxes and water movement in soil, SECHIBA must receive a number of input data from the atmosphere. They come either from observation data in a point or a region or from a general circulation model. Atmospheric information can only come from meteorological data which are often a combination of observations and modeling results. The data set is called an atmospheric forcing and the simulation mode is called forcing offline, which is imposed on the model simulation. However no feedback from the surface to the atmosphere is possible.

The relief of the surface is not reproduced in this model but is taken into account implicitly in the variability of atmospheric forcing or in the general circulation model (GCM). Thus, only the vegetation has an impact on the turbulence near the surface.

Variable	Description	Unit
T_a	2-meters air temperature	K
q_a	2-meters air humidity	kg.kg ⁻¹
W_N	Wind speed at 10 meters (u)	m.s ⁻¹
W_E	Wind speed at 10 meters (v)	m.s ⁻¹
P_{surf}	Surface pressure	Pa
SW_{down}	Short wave Incident Radiation (sun radiation)	W.m ⁻²
LW_{down}	Long wave Incident Radiation (infrared radiation)	W.m ⁻²
P_{liq}	Rain	kg.m ⁻² .s ⁻¹
P_{snow}	Snow	kg.m ⁻² .s ⁻¹

Table 2. 3 Input Variables received by SECHIBA

Variables forcing SECHIBA are summarized in Table 2.3. In forcing mode, the air temperature and humidity are generally given at 2 meters and the wind at 10 meters. Corrections, especially for the wind speed, must be applied to compute a correct friction coefficient and turbulent fluxes.

2.2.2 Energy Budget

The dynamic of the fluxes modeled in the energy budget are presented in Fig 2.1. Fluxes equations and descriptions are summarized in Table 2.4.

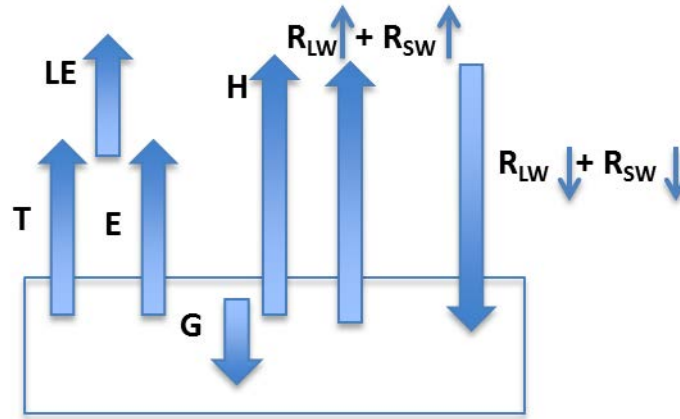


Figure 2. 1 Energy Balance

The energy budget main fluxes are part of the energy equation, which has the following form

$$R_n = LE + H + G \quad (2.3)$$

where R_n is the net radiation on the surface, LE represents the latent heat flux or instantaneous

energy consumed by the evaporation from the surface or received by condensation on the surface, H corresponds to sensible heat flux or energy (received or dissipated to the surface) exchanged by convection between the surface and the air and G is the exchanged heat flux between soil surface and depth. In Fig.2.1 T corresponds to transpiration, E is the evaporation and $R_{lw} + R_{sw}$ correspond to the incoming and outgoing long wave and shortwave incident solar radiations. All these fluxes are expressed in $W.m^{-2}$.

Flux	Equation	Description
Radiation	$R_{net} = R_{net}^{SW} + R_{net}^{LW}$ (2.4)	R^{SW} is the shortwave incident radiation α^{SW} is the surface albedo k_{emis} is the surface emissivity multiplying factor to be optimized
	$R_{net}^{SW} = k_{albedo} (1 - \alpha^{SW}) R^{SW}$ (2.5)	R^{LW} is the thermal radiation R_{surf}^{LW} is the thermal radiation emitted by surface
	$R_{net}^{LW} = k_{emis} (R^{LW} - R_{surf}^{LW})$ (2.6)	k_{albedo} is a multiplying factor weighing the effect the α^{SW} has in the computation of R_{surf}^{SW} . This parameter is optimized
Soil heat flux	$\frac{\partial T}{\partial t} = \frac{k_{capa}}{k_{cond}} \frac{\lambda}{C} \frac{\partial^2 T}{\partial z^2}$ (2.7)	λ is the soil conductivity, C the soil heat capacity and T the soil temperature. The soil is discretized on 7 thermal layers. The layers have constant depths so that the first layer has a characteristic time of 30 minutes and the last of 2 hours. k_{capa} and k_{cond} are multiplying factors weighing the parameters λ and C . They are both part of the control parameter set
Turbulent Fluxes	Aerodynamic Resistance	
	$C_d = \left(\frac{\kappa}{\log_{10} (k_{z0} \cdot z_0)} \right)$ $r_a = \frac{1}{\ U\ C_d Z_{0_{overheight}}}$ (2.8)	C_d is the drag coefficient. $\kappa = 0.41$ is the Von Karman Constant z_0 the roughness length and k_{z0} is a multiplying factor to be optimized. U is the normalized wind ($m.s^{-1}$),
Sensible heat flux	$H = \frac{\rho}{r_a} c_p (T_s - T_a)$ (2.9)	H is proportional to the temperature gradient in the surface-atmosphere interface. ρ the air density ($kg.m^{-3}$), C_p the air heat capacity ($J.kg^{-1}.K^{-1}$) and T_a and T_s the air and surface temperature in Kelvin
Latent Heat Flux	$LE = \rho \ U\ C_d \beta (q_{air} - q_{sol})$ (2.10)	
	Bare Soil Evaporation $r_l = h_s \cdot rsol_{cste}$	
	$E = \frac{1}{1 + \frac{r_l}{r_a}} U_s \cdot E'_{pot}$ (2.11)	LE integrates evaporation and transpiration. β corresponds to a coefficient integrating evaporation and transpiration resistance coefficients. In SECHIBA, LE is divided in three components: Bare soil evaporation, interception loss and transpiration E'_{pot} is the potential evaporation, U_s the soil humidity computed in the hydrological balance, r_l is the bare soil evaporation resistance, $rsol_{cste}$ equal to $33000 s.m^{-2}$, representing the resistance for bare soil square meter. This parameter is taken in the control parameter set. h_s is dry soil height.

Interception loss	$E_k = \min \left(I_k, \frac{I_k}{I_k^{\max}} \cdot \frac{1}{1 + \frac{r_k}{r_a}} \cdot E'_{pot} \right) \quad (2.12)$	I_k is the foliage intercepted water and I_k^{\max} is the maximum quantity that can be intercepted ($I_k^{\max} = LAI \cdot 0.1 \text{ mm}$), r_k the vegetation structural resistance, given by Table 2.2
Vegetation Stomata Resistance	$r_k^v = k_{rveg} \left(\frac{1}{LAI_k} \frac{R_{SW}^n + R_0}{R_{SW}^n} \frac{a + \lambda \delta c}{k_0} \right) \quad (2.13)$	k_{rveg} is a multiplying factor, weighing the calculation of the vegetation resistance to transpiration. This parameter is optimized. $R_0 = 12.5 \text{ W.m}^{-2}$, $a = 2.3 \cdot 10^{-2} \text{ kg.m}^{-3}$ and $\lambda = 1.5$. R_{SW}^n is the net solar radiation: $R_{SW}^n = (1 - \alpha_{SW}) R_{SW}$. δc is the atmosphere water vapor deficit: $\delta c = \rho \max(q_s(T_a) - q_a, 0)$.
Transpiration	$T_k = f_k \rho \ U\ C_d \beta_3 (q_{sol}^{sat} - q_{air}) \quad (2.14)$	Transpiration is computed for each vegetation fraction. β_3 is a coefficient equivalent to the vegetation stomata resistance

Table 2. 4 Energy budget fluxes

2.2.3 Hydrological Budget

The SECHIBA version used in this work models the hydrological budget based on a two-layer soil profile (Choisnel model, 1977). The soil layers correspond to the surface and the bottom of the soil. The total depth of the layers corresponds to the plants root depth. The soil has a unique type, with a total depth of $dpu_{cste} = 2m$. The bottom layer of soil acts like a bucket that fills with water from the top layer. When the top layer is empty with no water (due to evaporation, drainage to the lower layer, or lack of precipitation), this layer disappears. When rainfall exceeds the evaporation losses, they recreate a wet surface, allowing it to evaporate. If the water quantity is about to saturate the two soil layers, the top layer disappears again and excess water is removed by runoff.

The soil fluxes, as they are modeled in SECHIBA, are presented in Fig 2.2. The different operations to obtain the fluxes are summarized in Table 2.3.

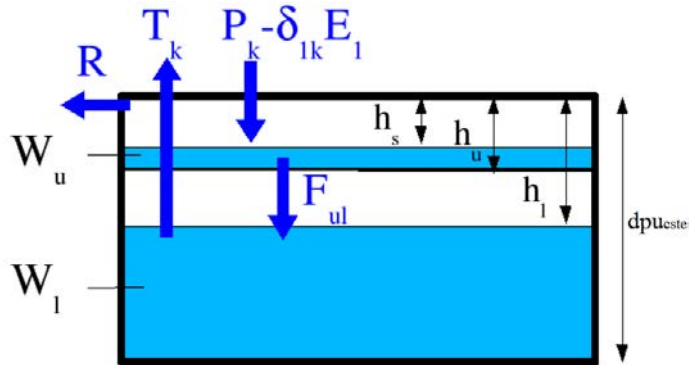


Figure 2. 2 Specific variables involved in hydrological budget computing (d'Orgeval, 2006).

In Fig.2.2, W_u is the water content in the top layer, W_l is the water content in the bottom

layer, h_s is the height of the dry part of soil in the top layer, h_u is the height of the top layer and h_l includes h_u plus the height of the dry part of soil in the bottom layer.

Flux	Equation	Description
Atmospheric Flux	Incoming precipitation $P_k = \max(P - (I_k^{\max} - I_k), 0)$ $I_k = \min(I_k^{\max}, P + I_k)$ (2.15)	The sub index k corresponds to the vegetation type. Once P_k is estimated, soil reservoirs are updated with the atmospheric fluxes.
	$\frac{\partial w_l^k}{\partial t} = -T_k$ $\frac{w_u^k}{\partial t} = P_k - \partial_{lk} E_l$ (2.16)	Bare soil evaporation is subtracted only from the bare soil vegetation type (δ_{lk} is equal to 1 if $k=l$ and zero if not, and $T_l = 0$) Interception reservoirs are updated extracting the evaporation
	Maximum water content in the soil $W_{\max} = dpu_{este} \cdot mx_{eau}$ (2.17)	mx_{eau} is an internal parameter of the model and its equal to 30 kg.m^{-3} . dpu_{este} is total soil depth. Both are optimized.
2-layer water content update	$R = W_u + W_l - W_{\max}$ $W_l = W_{\max} \text{ and } h_l = 0$ $W_u = 0 \text{ and } h_u = h_s = 0$ (2.18)	if $W_u + W_l \geq W_{\max}$, the water goes as runoff, and the surface layer disappears For more references in the computing of this flux, refer to d'Orgeval (2006), Kuppel (2012). □
Flux between layers	$(1) F_{ul} = F_{\min} \frac{W_u}{h_u \cdot mx_{eau}}$ $(2) F_{ul} = F_{\min} \frac{W_u}{h_u \cdot mx_{eau}} +$ $(F_{\max} - F_{\min}) \left(\frac{W_u - W_{\lim}}{h_u \cdot mx_{eau} - W_{\lim}} \right)^b$ (2.19)	F_{ul} represents water flux between the 2 layers. If $W_u < \frac{3}{4} h_u mx_{eau}$, we have (1) if not, we have (2). $F_{\min} = 2.10-3 \text{ mm.h}^{-1}$ and $F_{\max} = 0.2 \text{ mm.h}^{-1}$
	$W_l = W_l + F_{ul} \text{ and } h_l = h_l + \frac{F_{ul}}{mx_{eau}}$ $W_u = W_u - F_{ul} \text{ and } h_u = h_u - \frac{F_{ul}}{mx_{eau}}$ (2.20)	Upper and lower reservoir updates, with updates of the layer height.
Root extraction potential	$1) U_s = \exp(-hum_{este} \cdot h_l)$ $2) U_l = \exp(-hum_{este} \cdot h_l)$ $U_u = \exp\left(-hum_{este} \cdot dpu_{este} \frac{h_s}{h_u}\right)$ $U_s = \max(U_l, U_u)$ (2.21)	The root extraction potential U_s is computed for each vegetation type. The result is used in the energy balance computing on the next time step. The wilting point is defined as $W_{wilt} = 5 \text{ kg.m}^{-3}$ If $W_u + W_l < W_{wilt}$ then $U_s = 0$. If $h_u = 0$, there is no surface reservoir and we have case (1), contrary, if $h_u > 0$, we have case (2)

Table 2. 5 Hydrological budget fluxes

2.2.4. SECHIBA Parameters

Several key parameters in SECHIBA are considered in this work. Some parameters are standard internal parameters and others are multiplying factors which allow us to understand the importance of the several fluxes they are weighing. This subset of parameters are

presented in Table 2.6.

Parameter	Description	Prior Value	Unit	Equation
Internal Parameters				
hum_{cste}	Root extraction potential coefficient	Refer to Table 2.2	m^{-1}	2.21
$rsol_{cste}$	Evaporation resistance	33000	S/m^2	2.11
min_{drain}	Diffusion between reservoirs	0,001	-	
dpu_{cste}	Total depth of soil water pool	2	m	2.17, 2.21
mx_{eau}	Maximum water content	150	Kg/m^3	2.17, 2.19, 2.20, 2.21
Multiplying Factors				
k_{emis}	Surface Emissivity	1	-	2.6
k_{capa}	Soil Capacity	1	-	2.7
k_{cond}	Soil Conductivity	1	-	2.7
k_{rveg}	Vegetation Resistance	1	-	2.13
k_{z0}	Roughness height	1	-	2.8
k_{albedo}	Surface albedo	1	-	2.5

Table 2. 6 SECHIBA Parameters studied in this work. There are 6 internal parameters, involved in the model estimations and 5 multiplying factors that are imposed to specific fluxes

Model internal parameters (Table 2.6) are the following: $rsol_{cste}$ is a numerical constant involved in the soil resistance to the evaporation: this parameter limits the soil evaporation so the greater its value the lower the evaporation; hum_{cste} and mx_{eau} are related to soil water processes: the higher their values, the more water will be available in the model, affecting water transfers and especially evapotranspiration; dpu_{cste} represents the soil depth in meters. The other parameters are multiplicative factors. We consider k_{rveg} which is used in the calculation of the stomatal resistance, this variable limits the transpiration capacity of leaves, the greater its value, the lower the transpiration; k_{emis} is the soil emissivity used to compute soil surface temperature. This parameter takes part in the net radiation calculation which determines the balance between incoming and outgoing energy at the surface; k_{albedo} weighs the surface albedo which is defined as the reflection coefficient for short wave radiation; k_{cond} and k_{capa} take part in the thermal soil capacity and conductivity, both involved in the computation of the soil thermodynamics and k_{z0} weighs the roughness height which determines the surface turbulent fluxes.

2.3 Data

In this section we present the data used to constrain the land surface model SECHIBA. Data is the primary source of information for the understanding of a natural process. It serves not only to the construction of adequate equations, but also to validate the model outputs. The data used in this work correspond to *in situ* measurements of the forcing necessary to model the energy balance and water balance with SECHIBA. They constitute direct observation, quasi-continuous with high temporal resolution (30 minutes).

Several types of sensors and methods have been implemented to improve the observability of the biosphere processes. The data retrieved, in a context of data assimilation, is an indispensable source of understanding, modeling, validation, and improvement of the modeled process. Three *in situ* sites were selected for this study on SECHIBA YAO modeling. They are presented in the next section.

2.3.1. Eddy Covariance measurements

Principle

Flux measurements are widely used to estimate the exchange of heat, water, and carbon dioxide, as well as methane and other trace gases. The eddy covariance method is one of the most direct ways of measuring such fluxes. In our work, measurements from FLUXNET Network are used, which implement eddy covariance methods to measure the exchange of water vapor, carbon dioxide, among others fluxes between terrestrial ecosystems and the atmosphere.

The eddy covariance method (Aubinet et al. 2012) (also known as eddy correlation and eddy flux) is used to measure and calculate vertical turbulent fluxes densities of CO₂ (F_c), latent (LE) and sensible heat (H) flux within vegetation and atmospheric layers. The measured variables are compared to the mean covariance between vertical wind velocity (w') and scalar (c') fluctuations (As presented in Figure 3.4). In addition, positive flux densities represent mass and energy transfer into the atmosphere and away from the surface (Baldocchi et al. 1988). Turbulent fluctuations are computed as the difference between instantaneous and mean scalar quantities.

In our sites, eddy covariance is used mainly to determine gas emissions rates from land and water areas. It can also be used to estimate carbon dioxide, methane and water vapor fluxes.

The description of the eddy covariance method is based on the work of Burba, 2013. Air flow can be imagined as a horizontal flow with numerous rotating eddies (as shown in Fig.2.3). Each eddy has 3-D components, including vertical motion of the air.



Figure 2. 3 Representation of horizontal air flow that passes through the tower, and consists of differently sized eddies. (Source Burba, 2013)

Closer to the ground, smaller eddies rotate faster, and hence, there is an increased transport by higher frequency movements of air. Further away from the ground, larger eddies rotate slower, and hence, the transport is increased by lower frequency movements of air.

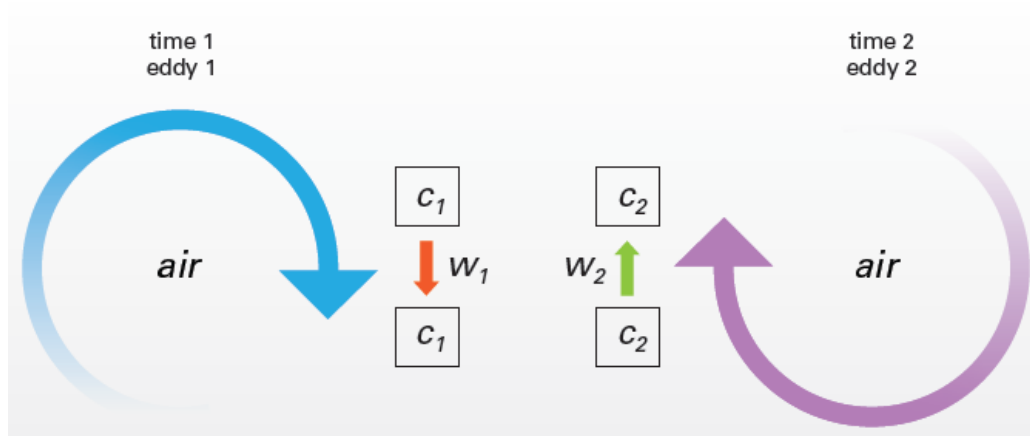


Figure 2. 4 The air flow consist of rotating eddies. Eddy 1 moves parcel of air c_1 down with the speed w_1 , while eddy 2 moves parcel c_2 up with the speed w_2 . (Source Burba, 2013).

Each parcel c_i of air has a concentration, a temperature, and a humidity unit. At a specific point of a tower, by measuring these characteristics and the speed of the vertical air movement, we know the vertical upward or downward fluxes of gas and water vapor concentrations, temperature, and humidity.

FLUXNET Network

FLUXNET (Baldocchi et al., 2001) is a network containing regional and global analysis of measurements from different tower sites, spread around the globe (Fig. 2.5), and grouped in regional networks. The data from all networks is accessible to the scientific community via the Fluxnet website (<http://fluxnet.ornl.gov>).

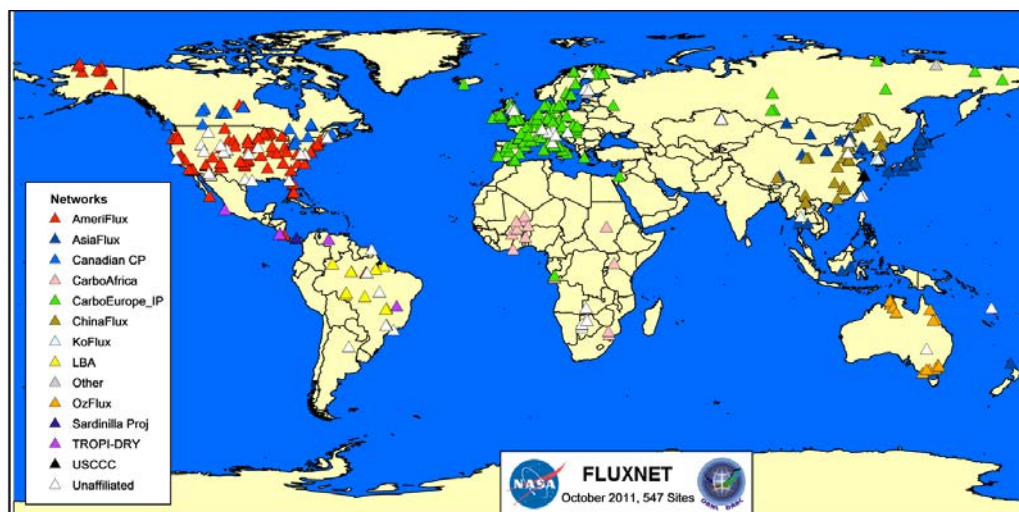


Figure 2. 5 Distribution of tower sites (source <http://fluxnet.ornl.gov>)

In this work, we selected two sites: Harvard Forest and Skukuza Kruger National Park. Both present contrasted climates and land surface properties to test the tools developed and to assess model parameter sensitivities. Only climate measurements (model forcing) from both sites are considered. Vegetation characteristics are prescribed and only homogeneous grids are considered. Two cases were studied with grassland (PFT 11) and bare soil (PFT 0). Forcing for both sites correspond to SECHIBA forcing, with the same sampling frequency (30 minutes).

Harvard Forest

Located in the United States of America, on land owned by Harvard University, the station is located at latitude 42°53'78" N and longitude 72°17'15" W. It was established in 1991. The site has a Temperate-Continental climate with hot or warm summers and cold winters. The annual mean precipitation is 1071 mm, the mean annual temperature is 6.62 °C and the altitude is 340 m.

In this work, we performed a simulation from 1992 to 1995 with the Harvard Forest forcing. The model state at the end of 1995 was retained in order to be used for our different experiments, which were made with 1996 forcing.

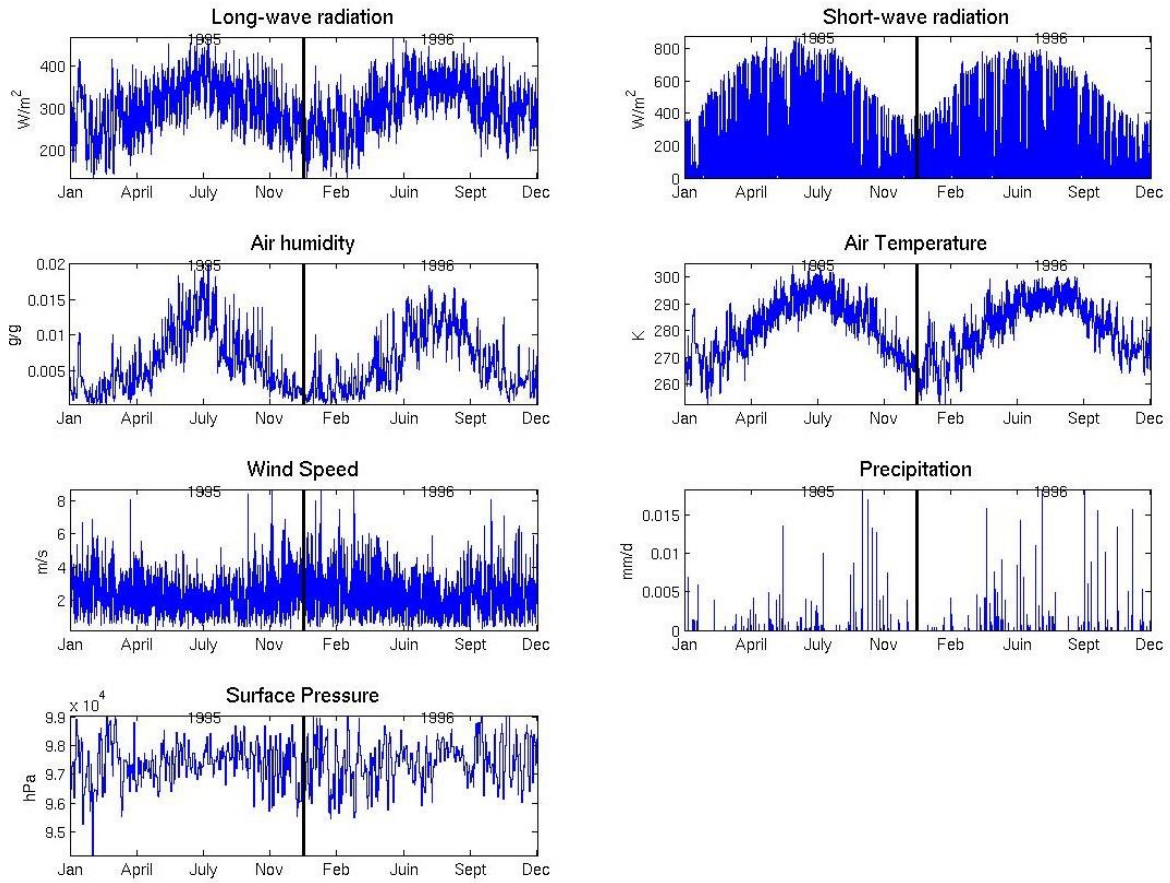


Figure 2. 6 Time series of forcing variables for FLUXNET Harvard Forest from 1995 to 1996.

Meteorological measurements from the last two years of data available are presented in Fig. 2.6. These forcing measurements correspond to ORCHIDEE inputs, in unit and scales. These atmospheric variables measured at the site impose local forcing over our surface model.

Skukuza Kruger National Park

Located at latitude $25^{\circ} 1' 11''$ S and longitude $31^{\circ} 29' 48''$ E, in South Africa, this Fluxnet site was established in 2000. The tower overlaps two distinct savanna types and collects information about land-atmosphere interactions. The climate is Subtropical-Mediterranean. The total mean annual precipitation is 650 mm, with an altitude of 150 m and the mean annual temperature is 22.15°C .

Data for Kruger Park site was available from 2001 to 2003. The first two years were used to calibrate the model and 2003 was used to perform the different experiments presented in this work.

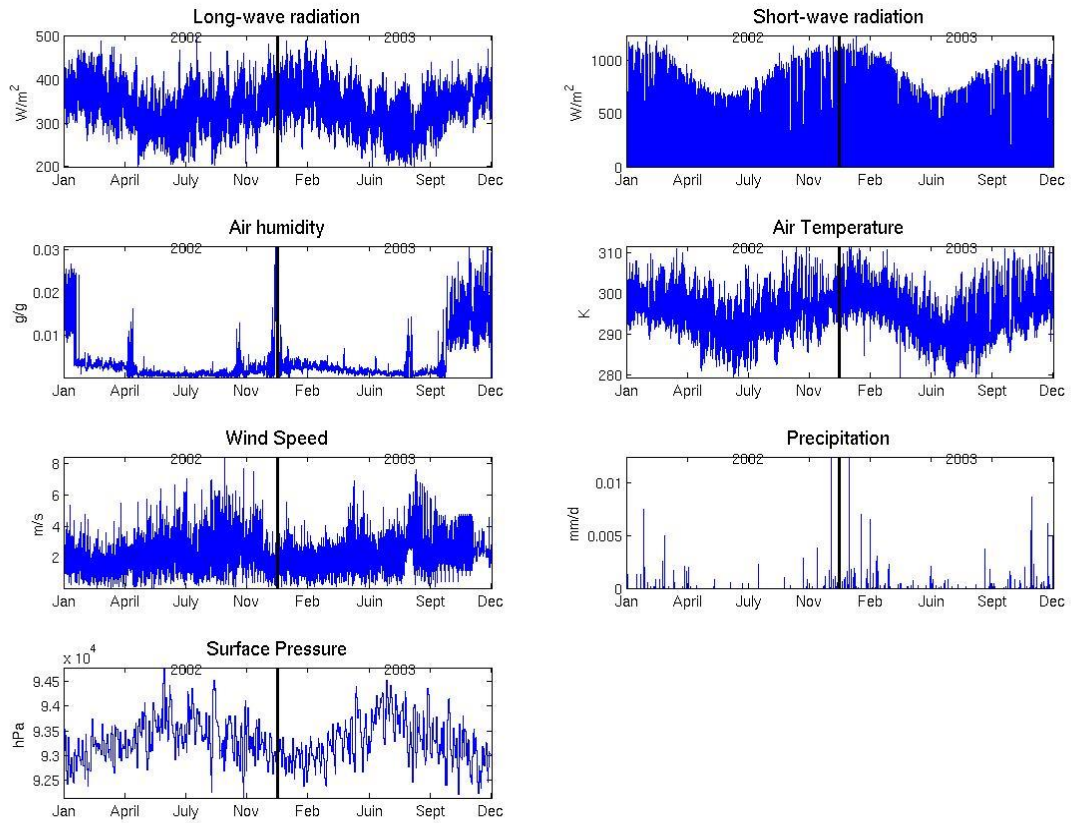


Figure 2. 7. Time series of forcing variables for FLUXNET Kruger Park from 2002 to 2003.

Meteorological measurements from the last two years of data available are presented in Fig. 2.7 for Kruger Park. Since Kruger Park is in South Africa in the southern hemisphere, annual cycles of fluxes are the inverse of those found in Harvard Forest site, which is located in the northern hemisphere. As in Harvard Forest site, forcing measurements shown in the curves correspond to ORCHIDEE inputs, in unit and scales. These atmospheric variables measured at the site impose local forcing over our surface model

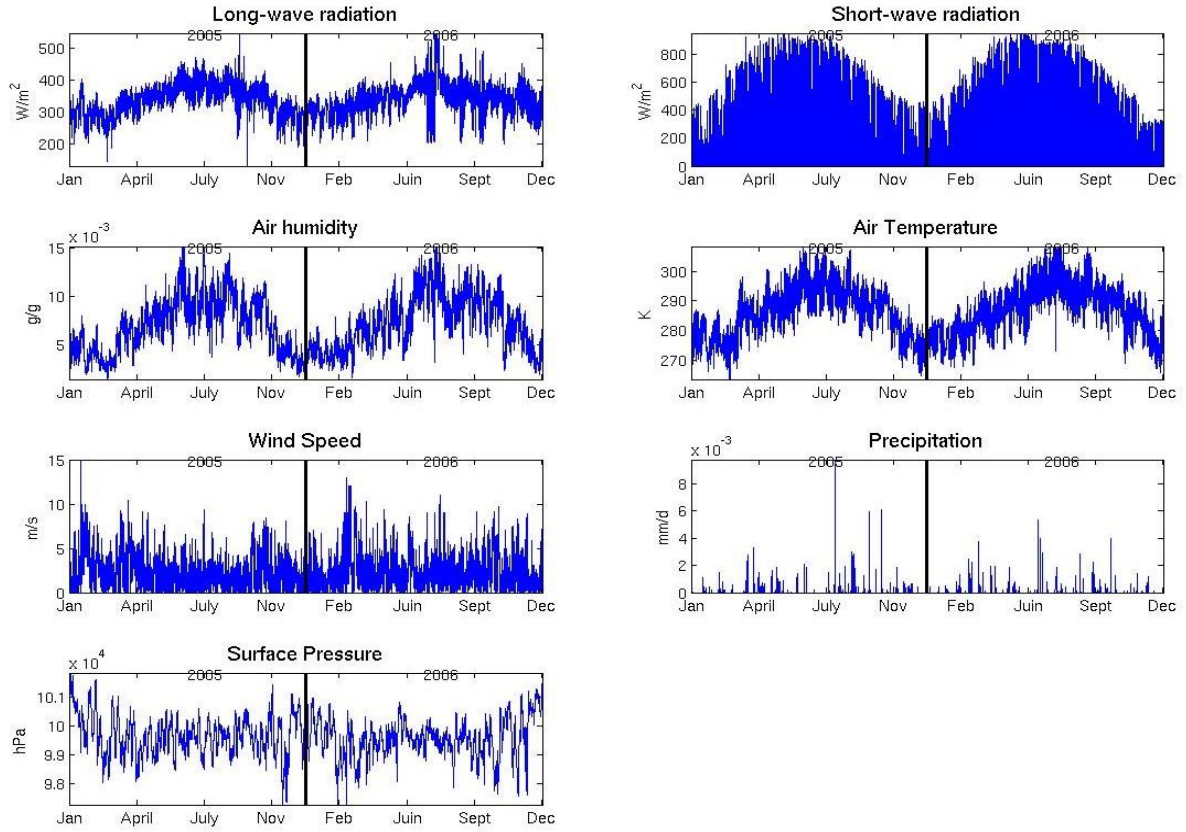
2.3.2 SMOSREX

SMOSREX is an experimental site situated on the complex of the ONERA (43°23'N, 1° 17'E, altitude 188 m above sea level), about 30 km south of Toulouse. At this site, a weather station continuously takes measurements of precipitation, temperature at 2 meters, surface fluxes, solar and infrared radiation, speed and wind direction, atmospheric pressure and specific humidity of the air.

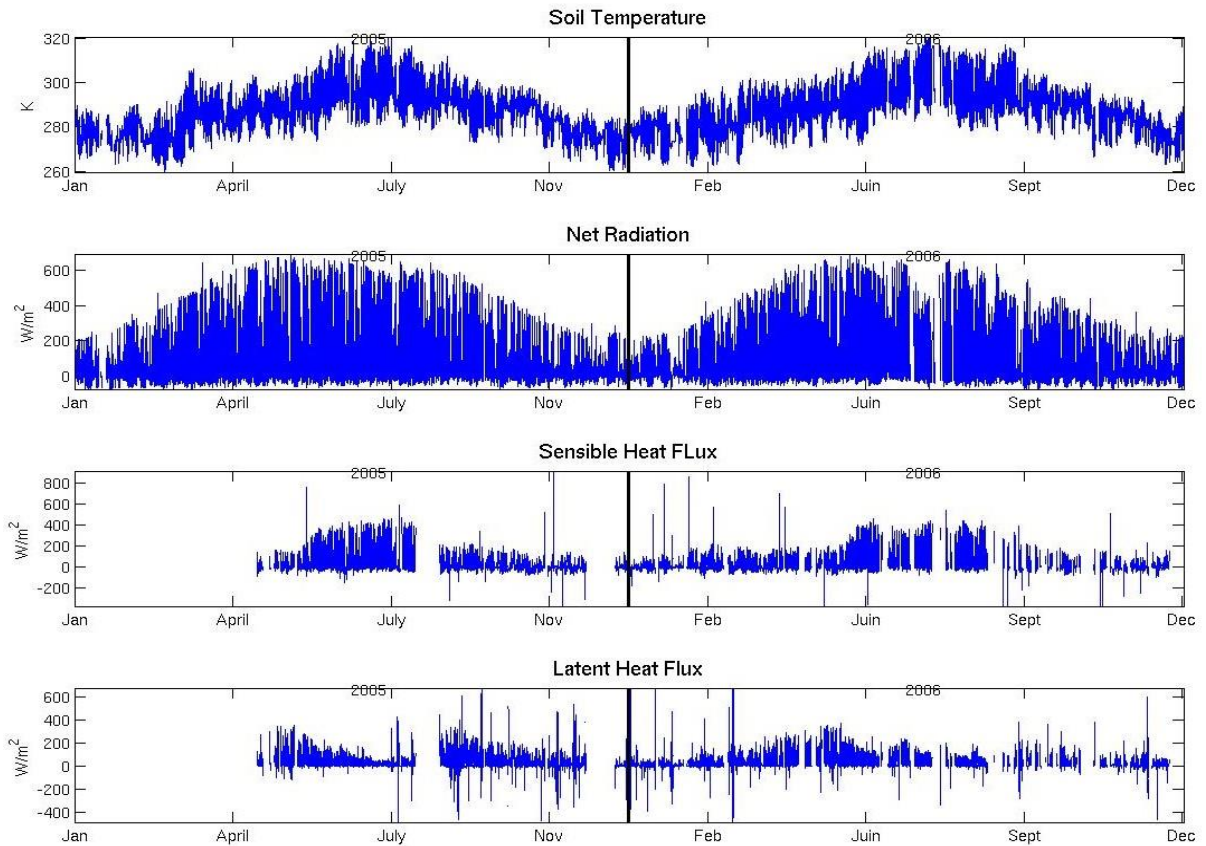
Observations and remote-sensing measurements were performed on two types of surfaces: bare soil and fallow (de Rosnay et al, 2006.). Weather conditions are very different with a cold wet winter and dry hot summer. The humidity and temperature in the soil profiles are characterized using measurements at different depths on two parcels of fallow and bare soil.

In this sites, the weather station measures atmospheric variables each 30 minutes, matching the frequency used in ECHIBA. Available measurements spread from 2003 to 2011. In this work we only used measurements from 2005 to 2006. These measurements are presented in Figure 3.9. As it can be seen, more precipitation periods are found in 2006, although in 2005 there are several periods during summer where the intensity of precipitation is high. Air temperature has important amplitude from day to night during the whole analyzed period. The inter-annual variability is significant in terms of characteristics of the annual cycle.

In addition to the climate measurements, observations of latent and sensible heat fluxes, soil temperatures and net radiation are available. However, fluxes have many missing data, as it can be seen in Fig. 2.8.(b), where there is no measurement of the latent and sensible heat fluxes in winter 2005. Measurements are more complete for soil temperature and net radiation, but there are still missing data as well. If one of these flux is considered as part of an assimilation process, these missing data have to be taken into account; no parameters restitution can be done during the missing time measurements.



(a)



(b)

Figure 2. 8 Time series of forcing variables (a) and measured fluxes (b) for SMOSREX, from 2005 to 2006.

Chapter 3

Theoretical Principles of Variational Data Assimilation

3.1 Introduction and Notation

The objective of this work, as was mentioned in Chapter 1, is to improve model state variables through the implementation of 4DVAR assimilation into SECHIBA, constraining model internal parameters. Once the model and the parameters to be optimized were presented in Chapter 2, this Chapter introduces the mathematical concepts necessary for the implementation of variational data assimilation (VDA).

Traditionally, VDA is classified into different types, depending on the number of dimensions: 3DVAR means that we consider a physical phenomenon described in space by one, two or three dimensions. In 4DVAR its evolution over time is also observed. It requires the knowledge of a numerical model or direct model M , describing the time evolution of the physical phenomenon that is being studied. If we take for example a geophysical problem, the direct model allows, among others, to link the studied geophysical variables to observations. By varying some geophysical parameters (control variables) assimilation seeks to infer the model physical variables that led to the observation values. These physical variables can be, for example, initial conditions or poorly known parameters of M . For this second case, the variations are considered from a set of initial values given to the control variables. In order to infer the right values of the control parameters, VDA aims to determine the minimum of a cost function J that measures the difference between the observations and their equivalent values calculated by the model M .

The desired minimum of the cost function is obtained from the gradient method. According to the algorithm considered, it is necessary to use models that are derived from M : the adjoint and the tangent linear model. If M is continuous and differentiable, these two models are derived from the equations of the model M . With the tangent linear model we study, for a specific control variable, the sensitivity of the output values of the model M to small perturbations of these variables at a reference point. It corresponds to the value of the Jacobian matrix at this point. With the adjoint model, we study changes in control variables, in response to a perturbation of the output of the model M . It must therefore be computed in the opposite direction to the linear tangent calculation, which implies using the transpose of the Jacobian matrix.

When observations are available, these models are used to implement VDA, minimizing the function J , and finding the values of the control variables. We now present the formal mathematical notation and the most conventional algorithm currently used in VDA. In this work, we adopted the formalism and notations presented in Nardi et al., 2009 and Thiria et al., 2006. We denote:

- M as the direct model describing the evolution between two discretization time steps t_i and t_{i+1} .
- $\mathbf{x}(t_0)$ represents the initial state vector of the model, which has to be controlled. It is assumed that it has dimension N .
- $M_i(\mathbf{x}(t_0))$ or $M(t_0, t_i)$ is the model state at time t_i from the model state at t_0 . We denote $\mathbf{x}(t_i) = M_i(\mathbf{x}(t_0))$.
- $\mathbf{M}(t_i, t_{i+1})$ means the tangent linear model, which is the Jacobian matrix of the model M calculated at $\mathbf{x}(t_i)$.
- \mathbf{x}^b denotes a background vector, which corresponds to a prior estimate of the vector $\mathbf{x}(t_0)$.
- \mathbf{y}^o represents the available observations. \mathbf{y}_i^o vector corresponds to observations at time t_i , this vector can be empty if there is no observation in this time step.

The tangent linear model of M_i calculated at $\mathbf{x}(t_0)$ is defined as

$$\mathbf{M}_i(\mathbf{x}(t_0)) = \prod_{j=i-1}^0 \mathbf{M}(t_j, t_{j+1}) \quad (3.1)$$

The adjoint model calculated in $\mathbf{x}(t_0)$ is the transposed matrix of the linear tangent, computed as follows

$$\mathbf{M}_i^T(\mathbf{x}(t_o)) = \prod_{j=0}^{i-1} \mathbf{M}(t_j, t_{j+1})^T \quad (3.2)$$

The direct model M is used to estimate state variables that can be matched to real measurements. This comparison is performed from an operator H called observation operator. H allows the transformation of the output values of M in observable values. The observations are not available at every point of the studied area or at every time step, consequently where observations are available, H links the observed values and those calculated by the composition $H \circ M$. In the field of geophysics, this operator allows, for example, to compare the outputs of the model M , which calculates the temperature at the surface of the sea, to observations recorded by a satellite.

The cost function J will be defined based on the available observations. It is necessary to express the tangent linear and the adjoint of the operator H . We denote:

- H_i as the observation operator which allows us to calculate the observation variables \mathbf{y}_i at time t_i from the state vector $\mathbf{x}(t_i)$. Subsequently, it is assumed that:
 $\mathbf{y}_i^o = H_i(M_i(\mathbf{x}(t_o))) + \varepsilon_i$. (ε_i is a random variable with zero mean). Thereafter, ε_i is called the observation error.
- \mathbf{H}_i represents the tangent linear model matrix of the operator H_i calculated at $\mathbf{x}(t_i)$.

Finally, the most general form of the cost function is defined as follows:

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{x}_0 - \mathbf{x}^b) \mathbf{B}_0^{-1}(\mathbf{x}_0 - \mathbf{x}^b) + \frac{1}{2} \sum_{i=0}^n (\mathbf{y}_i - \mathbf{y}_i^o) \mathbf{R}_0^{-1}(\mathbf{y}_i - \mathbf{y}_i^o) \quad (3.3)$$

The first part of the cost function represents the discrepancy to \mathbf{x}^b . The second part represents the distance between the observations and the model estimates. \mathbf{B}_0 is the covariance error matrix of \mathbf{x}^b and \mathbf{R}_0 is the covariance error matrix of \mathbf{y}^o , at time t_i , which is supposed to be constant over time, and the number of observations is represented by n .

The minimization of the cost function is made using the adjoint of the model. The procedure to implement the minimization is performed by choosing an optimization algorithm among the set of those suggested by the optimization techniques. Minimizers M1QN3 and M2QN1 (Gilbert and Lemaréchal, 1989), developed by INRIA are widely used for the VDA in the field of geophysics. It is also possible, in order to overcome some problems with the

convergence and efficiency, using an approximate gradient descent method, called incremental algorithm (Weaver et al., 2005, Real et al., 1994, Louvel 2001).

3.2 Adjoint Method

In order to minimize J , it is necessary to compute efficiently its gradient, with respect to the control parameters. In general, J is a scalar function that measures the difference between the outputs \mathbf{y} of a model (that we note as G) and observations, so J depends on \mathbf{y} . To minimize J , it is necessary to calculate its gradient with respect to the model variables \mathbf{x} as $\mathbf{y}=G(\mathbf{x})$. If we note $\nabla_{\mathbf{y}}J$ ($\nabla_{\mathbf{x}}J$ respectively) the gradient vectors of J with respect to \mathbf{y} (with respect to \mathbf{x} , respectively), we have

$$\nabla_{\mathbf{x}}J = \mathbf{G}_{\mathbf{x}}^T \nabla_{\mathbf{y}}J \quad (3.4)$$

This allows the computation of $\nabla_{\mathbf{x}}J$ knowing $\nabla_{\mathbf{y}}J$, in the form of a matrix product of this matrix by $\mathbf{G}_{\mathbf{x}}^T$, which is the transpose of the Jacobian matrix of G computed at \mathbf{x} . VDA consists then of minimizing the function J with respect to the initial state $\mathbf{x}(t_0)$ using the gradient method. Taking into account the notations introduced in Section 3.1, the development of the calculation of the gradient gives the expression

$$\nabla_{\mathbf{x}(t_0)}J = \mathbf{G}_{\mathbf{x}(t_0)}^T \nabla_{\mathbf{y}}J + \mathbf{B}^{-1}(\mathbf{x}(t_0) - \mathbf{x}^b) = \sum_{i=1}^n \mathbf{M}_i^T(\mathbf{x}(t_0)) \mathbf{H}_i^T [\mathbf{R}_i^{-1}(\mathbf{y}_i - \mathbf{y}_i^o)] + \mathbf{B}^{-1}(\mathbf{x}(t_0) - \mathbf{x}^b) \quad (3.5)$$

The control parameters are adjusted several times until a stopping criterion is reached. The iterations of the gradient method allow us to approach a local minimum of J , which gives a possible value of the initial condition $\mathbf{x}(t_0)$.

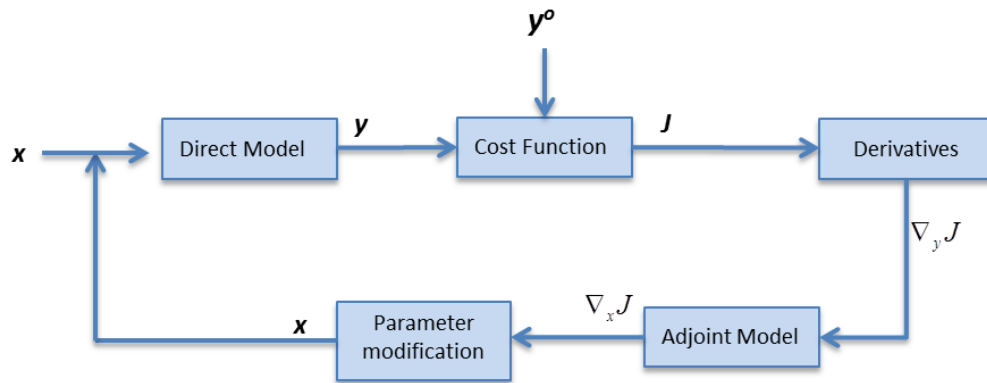


Figure 3. 1 Basic iteration of variational data assimilation. (Brajard, 2006)

Fig.3.1. shows a basic iteration of the VDA. \mathbf{x} is the vector $\mathbf{x}(t_0)$ in the first iteration. The difference between the output of the forward model \mathbf{y} and the observations \mathbf{y}^o is measured using the cost function J . As shown in Eq.(3.5) in order to minimize this function, it is necessary to use $\nabla_{\mathbf{y}} J$ through the adjoint model, giving as result $\nabla_{\mathbf{x}} J$.

The objective of this work is to show the capacity of 4DVAR to help determining the value of the principal parameters of SECHIBA, investigating the impact of the prior choices, thus as the principal part of the study concerns the direct model M and twin experiments, we take \mathbf{H} as the identity matrix and take no background. Regarding the error covariance matrices \mathbf{R}_o and \mathbf{B}_o , since this work is a prospective study, these correlations are not taken into account, and thus both matrices are identity matrices. With this simplification the cost function in the different assimilation experiments presented in this work is defined as follows

$$J(\mathbf{x}_0) = \frac{1}{2} \sum_{i=0}^n (\mathbf{y}_i - \mathbf{y}_i^o)(\mathbf{y}_i - \mathbf{y}_i^o)^T \quad (3.6)$$

The observations \mathbf{y}_i^o correspond to the land surface temperature measurements, which are used in the following experiments.

3.3. Representing a model and its adjoint through a modular graph

The description of this section is based on the work by Nardi *et al*, 2009. Considering the case of a discrete physical model (atmospheric, oceanic or other), the state evolution in space between two consecutive time steps is governed by a discrete model of evolution: the direct model M . As an example, we consider here the case of a model in two spatial dimensions. A state variable is denoted by $x_{ij}(t)$ where i, j are the spatial coordinates on a mesh and t the time index. The discretization leads to a numerical expression for each variable $x_{ij}(t)$, which gives its functional dependence on other variables that may be located on any points of coordinates. These dependencies result in a sequence of functions corresponding to a simple modular graph.

A modular graph is a set of interconnected modules. The modules represent the nodes of the graph. A directed arc from F_p module to F_q means that F_p transmits a portion of its output to F_q . We also define:

- Module: A module F is a computational entity; it receives an input vector and computes an output vector. A graph is formed by several modules; a module receives its inputs

from other modules or from the external environment, and transmits data at the output of the other modules or to the outside environment.

- Basic connection: we schematize the transmissions between modules with connections called basic connections. They link the i^{th} output of module F_p to the j^{th} input of F_q module, indicating that the i^{th} value calculated by F_p is transmitted to the j^{th} of F_q . Transmission of data outside or inside the graph will be represented by a basic connection starting from the outputs of some of the modules. A physical model M is represented by a modular graph, denoted by Γ .

In Fig.3.2 a subset of modules is represented as an interconnected graph. F_p module transmits the output y_{p1} to the input x_{q1} of module F_q and output y_{p2} to input x_{q2} . The same output y_{p2} will also be sent to input x_{i1} of module F_i ; equally for the other modules. Fig.3.2.a represents the basic connections describing the transmission of data between modules. Fig.3.2.b represents the modular graph. In general, an arc of a modular graph represents several basic connections.

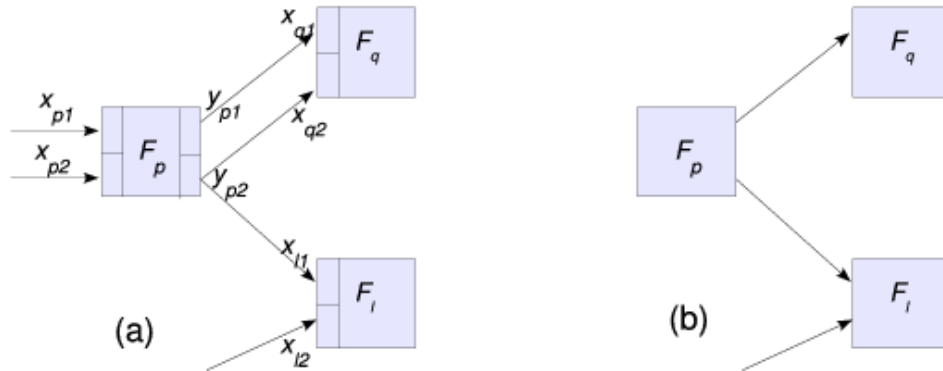


Figure 3. 2 Part of a modular graph with 3 modules. Only one basic connection is allowed by a given input, but multiple basic connections can emerge from the same output. Source: Nardi et al, 2009

The modular graph describes the ordering of calculation in a model (Fig.3.3). An arc from F_p to F_q indicates that the module F_q can trigger its calculation after that of F_p . The basic connections are used to transmit data.

The modular graph is a graph without circuits. It has at least one input vertex (without predecessors) and at least one vertex output (without successors), so there are three types of modules:

- The input modules (without predecessors) receive their data only from the external

environment of the graph and send their output to other modules or external environment.

- Output modules (without successors in the graph) transmit their outputs only to the exterior context of the graph. They receive data from other modules or from the exterior environment.
- The internal modules of the graph necessarily receive inputs from predecessors and possibly from the external environment and transmit the results to the successors and possibly the external environment.

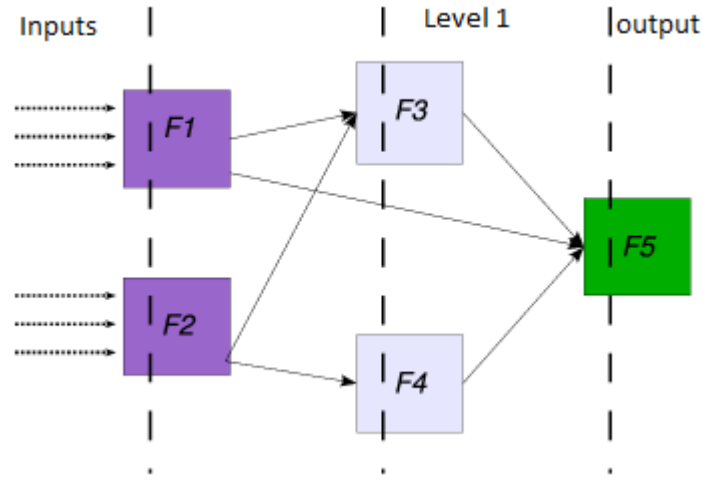


Figure 3. 3 Levels in a modular graph. F1 and F2 are input modules, F3 and F4 are internal modules and F5 is an output module. Source: Nardi et al, 2009.

As the modular graph is acyclic, it is possible to number the modules in an order called topological order. Thus, with respect to this order, the existence of an arc of F_p to F_q ($F_p \rightarrow F_q$) necessarily implies that $p < q$. All inputs of a module F_p are grouped in a vector noted \mathbf{x}_p . All of its output are integrated in a vector denoted \mathbf{y}_p ($\mathbf{y}_p = F_p(\mathbf{x}_p)$). From the above, a given module F_q can only be activated if it has its vector \mathbf{x}_q , which implies that all his predecessors have previously been activated. Inputs for basic connections, coming from outside, are initialized by the external environment. The topological order indicates a possible concatenation of the calculation in the model. The propagation of the calculation in the modular graph is done in a natural way using the following procedure, called the “*forward*” procedure:

(1) Forward algorithm

1) Browse modules of the modular graph going forward, in the direction of the increasing topological order.

For each module encountered:

- If it is an input module, initialize its inputs by the corresponding data from the external context.
- For other cases (output module or internal), if the module takes a part of its inputs from the external environment then initialize these inputs by the corresponding data from the external environment, initialize the other inputs by the corresponding outputs of the modules predecessors.
- Trigger the calculation of the module to obtain the results of its output values ($\mathbf{y}_p = F_p(\mathbf{x}_p)$)

2) Retrieve the result as a vector \mathbf{y} , consisting of data transmitted to the external environment

We show how, based on the concept of modular graph, we calculate the tangent linear and the adjoint of a model. The calculations require matrix multiplications involving the Jacobian matrices of each module. The procedure to obtain the tangent linear and the adjoint of a modular graph are presented in Appendix A.

3.3.1. Deployment of a modular graph

The physical model is computed at each time step and discrete grid point. The evolution of the model in the grid is similar at each grid point, having a repetition of calculation. The modular graph Γ , associated with the numerical model M , takes into account this repetition by:

- Defining a modular sub-graph (Γ_g), which describes the computations needed at time t for a given grid point (Fig. 3.4.a).
- The defined graph is a modular graph whose vertices are Γ_g and the arcs represent the exchanges between them (Fig. 3.4.b).
- The complete graph Γ , with its evolution through time, is obtained by duplicating the graph as long as necessary.

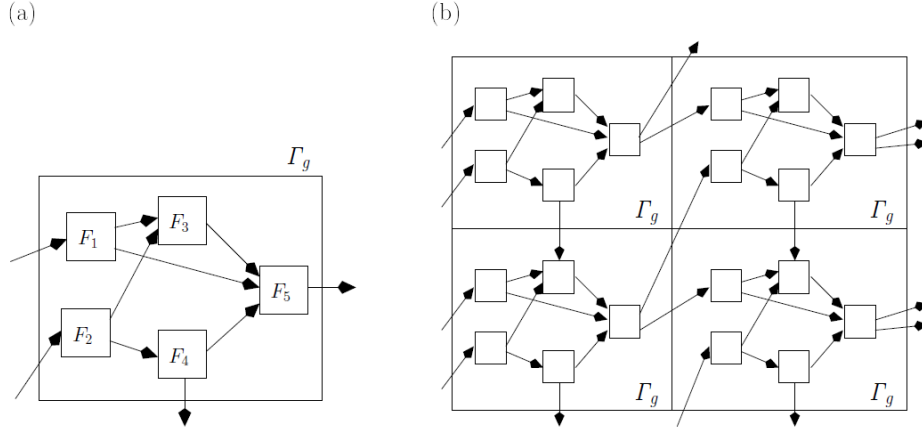


Figure 3. 4 Two graph abstraction levels. At the lower level (a), we build the graph Γ_g ; at the space level (b), the same graph Γ_g is repeated for each grid point (2D in this example). The space connections between the Γ_g graphs correspond to the basic connections between the modules. Source: Nardi et al, 2009

The basic connections coming from the external context of Γ could be, for example, initializations or boundary conditions. Outgoing basic connections transmit their values to compute, for example, a cost function.

3.4. Diagnostic tools for the assimilation system

An assimilation system is a complex machinery that requires an effective but robust functioning. This complexity leads to the definition of diagnostic tools in order to track several aspects of the system, to make sure the output is reliable. A number of diagnostic tools are presented in this section, listing the different parts of the assimilation system and how to test their outputs. More information about the diagnostics tools presented in this section can be found in Järvinen (1998) and Errico (1997).

3.4.1 Test the correctness of the adjoint model

A test for the correctness of the adjoint model is based on verifying the following scalar product

$$\forall dx, dy \quad \langle \mathbf{M}^T(x_0).dx, dy \rangle = \langle dx, \mathbf{M}^T(x_0).dy \rangle \quad (3.7)$$

With a machine precision tolerance depending on the relative error, computed with the following expression

$$\frac{\langle \mathbf{M}^T(x_0).dx, dy \rangle - \langle dx, \mathbf{M}^T(x_0).dy \rangle}{\langle \mathbf{M}^T(x_0).dx, dy \rangle} \quad (3.8)$$

The adjoint test provides no indication of the validity of the derived program; it simply indicates the consistency between the linear tangent model (*forward*) and the adjoint model (*backward*), which is already not negligible.

3.4.2. Test the correctness of the cost function gradients

Testing the gradient of the cost function J consist in the evaluation of the gradient, which must asymptotically point to the same direction as the difference between two realizations of the cost function which are separated by a small perturbation in model state (Järvinen, 1998). The test is derived by using first order or second order of Taylor expansion, as follows

	First Form	Second Form
1 st Order	$J(x_0 + \alpha \cdot dx) - J(x_0) - \langle \nabla J(x_0), \alpha \cdot dx \rangle \rightarrow 0, \quad (3.9)$ $\alpha \rightarrow 0$	$\frac{J(x_0 + \alpha \cdot dx) - J(x_0)}{\langle \nabla J(x_0), \alpha \cdot dx \rangle} \rightarrow 1 \quad (3.10)$ $\alpha \rightarrow 0$
2 th Order	$\frac{1}{\alpha^2} (J(x_0 + \alpha \cdot dx) - J(x_0) - \langle \nabla J(x_0), \alpha \cdot dx \rangle) \rightarrow C \quad (3.11)$ $\alpha \rightarrow 0$	$\frac{1}{\alpha} \left(\frac{J(x_0 + \alpha \cdot dx) - J(x_0)}{\langle \nabla J(x_0), \alpha \cdot dx \rangle} - 1 \right) \rightarrow K \quad (3.12)$ $\alpha \rightarrow 0$

In Equations 3.11 and 3.12, C and K are numerical constant proportional to the Hessian matrix. The test can be applied during developing time, on individual routines, but it can be also implemented to the whole model.

When this test fails, a general error in the variational assimilation system might be the cause. The error is not necessarily just in the adjoint coding. In order to explore the different possibilities to find the problem, a first approach consist in working with a reduced definition of the problem. If a coding error in the adjoint was made, the cost function gradient test may pass, given that errors in coding might produce only relatively small errors in the gradient computation. It is important to keep testing the adjoint codes as explained above.

3.4.3. Derivative test

The idea is to verify for each input of each module the following expression

$$\frac{F_j(x_1, \dots, x_i + \delta \cdot x_i, \dots, x_n) - F_j(x_1, \dots, x_n)}{\delta x_i} \approx_{\delta x_i \rightarrow 0} \frac{\partial F_j}{\partial x_i} \quad (3.13)$$

Where F_j is the j^{th} module of the modular graph and x_i its correspondent i^{th} input. In other words, the test is carried out for all the derivatives of the module, i.e. for each derivative

of each output with respect to each input (i.e. the Jacobian).

3.5. Summary

In this chapter, the general aspects of VDA were presented. It is important to be aware of the different components necessary to implement VDA, since they drive the assimilation process, allowing us to merge observations into the model. In addition the modular graph approach was presented, as an alternative representation of a numerical model. By decomposing the model into simpler routines, we can obtain a scalable and easily maintainable code, as well as a representation of the direct and adjoint model.

VDA and modular graph approaches are the base for our implementation methodology of 4DVAR, based on the use of a tool called YAO. Theoretical aspects of the tool and the implementation process of SECHIBA into YAO are presented in Chapter 4.

Chapter 4

The YAO Approach: Theoretical aspects and implementation of SECHIBA-YAO 1D

4.1 Introduction

In Chapter 3, the notions of variational data assimilation (VDA) were presented, as well as the representation of a numerical model based on a modular graph approach. At a computational level, two difficulties arise when implementing VDA: the first is to obtain a program to calculate the adjoint of the model; the second is the execution of VDA with a specific scenario.

In a VDA system, the user programs the discretized dynamical model, the gradient of the cost function through the adjoint model and the minimization operations following a specific set-up. From the data-processing point of view this involves two types of problems: First, if a direct model program already exists, it is necessary to implement the program which provides the adjoint model and its tangent linear model; second, once all these models have been implemented, it is necessary to list the calculations according to a certain scenario and to the chosen minimization method. The first problem leads to use automatic differentiation software and the second to design specific software. In this chapter, YAO is presented as a possible solution to cope with these two problems.

The purpose of SECHIBA implementation into YAO is to provide the scientific community with a code that can easily incorporate changes in its physics and that is able to run very quickly assimilation experiments, allowing the researchers to focus on issues related to the assimilation itself. This chapter is highly operational. It presents the different technical steps and shows the difficulties when implementing SECHIBA into YAO. The model in

question has already been encoded in Fortran, having 25.000 lines of code. Given that the thesis required the recoding of the model in YAO format, this work is an important part of the thesis.

4.2 YAO approach

YAO, developed at LOCEAN laboratory in France, is a tool for software development. It facilitates the implementation of the adjoint method, including features that allow it to act as a platform to launch data assimilation scenarios. YAO is not an automatic differentiator. It does not generate the adjoint code from a program implementing the direct model. Instead, it is based on a numerical representation of the model as a modular graph, as presented in Section 3.3. The graph contains the calculation flow described by the model. By integrating forward and backward the graph, we obtain a computational method equivalent to the adjoint of the model at a given point.

With YAO, the user specifies the type of discretization through a meta-language, the scenario, the specification of the direct model, and the derivatives of each routine (line-by-line). YAO generates the direct model, the tangent linear model and the adjoint model. It also allows the user to choose an implementation for the cost function J according to the specific scenario. Once the modular graph of a model M is defined, YAO creates a framework for developing pre-programmed underlying modular graph and its associated functions.

The components defined in a YAO implementation of a model are organized in the following steps:

1. Management of different data needed for the application (parameters, forcing, etc).
2. Choice of the cost function: incorporating (if available) matrices (or operators) variance-covariance of the background and observation errors. Some cost functions are directly available in YAO. It is also possible to integrate other methods (i.e. dual (Louvel, 2001) or quasi-static (Pires et al., 1996)) that will be able to communicate with the modular graph application.
3. Choice of the minimizer: with YAO it is possible presently to use 2 of the minimizers of INRIA: M1QN3 and M2QN1 (Gilbert and Lemaréchal, 1989). Further minimization algorithms can be interfaced.
4. Tools for verifying the accuracy of adjoint and tangent models: YAO facilitates these verifications, allowing access to various validity tests: adjoint test, cost function test,

among others (Section 3.5.).

4.3. Creating a project with YAO

The implementation of a numerical model in YAO starts with the definition of the modular graph describing the dynamics of the model. This is the duty of the user. Elementary processes and interconnections (dependencies) between modules are defined in order to catch the essence of the model, defining the basic connections. In addition, the user specifies a sequence of execution for the modular graph, corresponding to a topological order. Since the user must reformulate its direct model in the formalism of modular graph, the direct and adjoint model does not show up in the form of separate codes, but as the results of the application of the *forward* and *backward* modular graph algorithms, as mentioned before.

In Fig.4.1, a schematic representation of YAO is presented. The executable program of the model is made by YAO, based on the user specifications. The big orange square is the YAO compiler, which generates the modular graph and the program executable based on the *Modules* files, the *Description* file and the *Chapeau* file, all created by the user. The output is the executable, which is used to launch specific assimilation experiments, based on scenarios defined in the *Instruction* file.

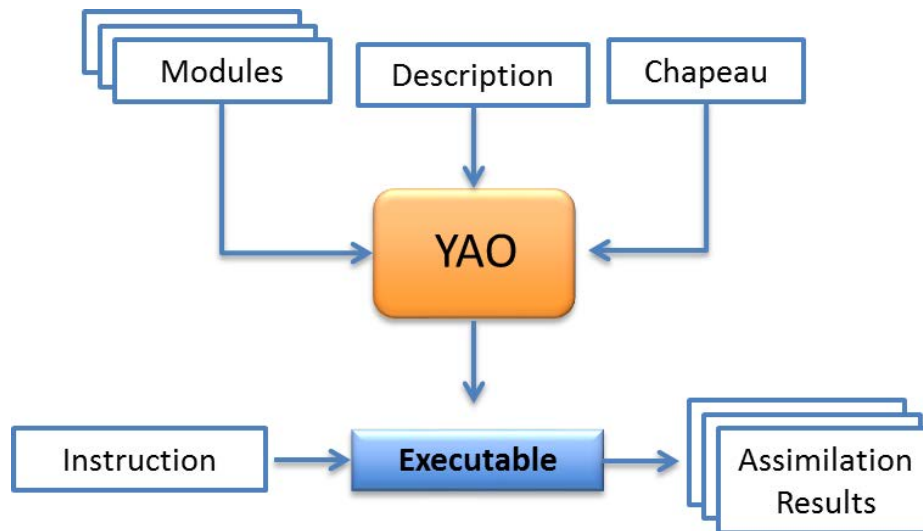


Figure 4. 1 YAO input and output components, from Nardi (2011). User specifies the *Modules*, *Description* and *Chapeau* files. YAO compiler generates the *Executable* and through the *Instruction* file, assimilations can be launched under specific scenarios

The *Description* file contains YAO guidelines, which define the modular graph of the direct model. Here we define: assimilation time window length, module definition, space dimension of each module, basic connections between modules, module computational order

and minimizer type.

The *Modules* files contain elements of the model as well as its Jacobian matrix. For each module, the user generates an independent file. They contain two functions: *forward* and *backward*; corresponding to the direct and the derivatives of the module, respectively

The *Chapeau* (Hat) file is the application main program. Here we define global variables and specific user functions. In addition there are a set of mandatory functions allowing the user to access in running time the calculations made at the beginning of the application and before or after the *forward* or *backward* computing.

The *Instruction* file is used to execute the application, passing specific user parameter and YAO options as well, using the YAO meta-language.

4.3.1 Input / Output Management

The management of the inputs and outputs in YAO depends on several user functions that facilitate the task of saving model states, loading observation, etc. As it can be seen in Fig.4.2, the user can upload observations and model states to the project environment, by using the function *loadstate*, *loadobs* and *outoobs*. These uploaded files must have a specific format allowing YAO to assign it to the correspondent modules.

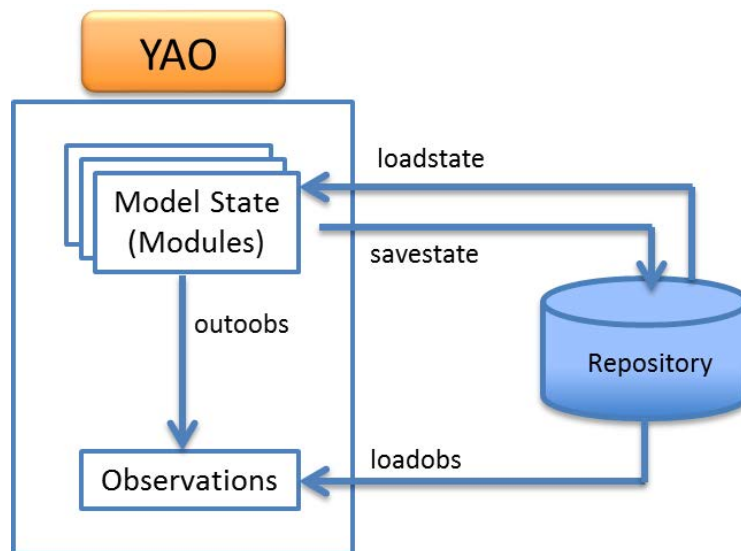


Figure 4. 2 Input and output data flow through YAO. Users can upload observations and model states to the project environment, following specific YAO notation.

Every observation and model state the user wants to upload must have its match in the modular graph. Observations must be declared with an extra keyword, identifying them as the

module used in the cost function. The function *savestate* allow the users to save the state of the entire model or a particular module in a specific time step or in an entire trajectory.

The save and load functions have a set of parameters, required to position in time and space the data flow. These parameters involve the module name, the output number, axis number, time step, coding type, and data format. More about the functioning of these functions can be found in the YAO documentation, by Nardi et al, 2009.

4.3.2 Diagnostic tools for the generated project

To verify that the model and its inverse have been properly implemented, there is a set of functions to assist the development provided by YAO. These four methods are briefly presented in this section. These test are explained in detail in Nardi (2011) and Kane (2010). These diagnostic tools are implementations of the test presented in Section 3.5.

The first test is *testdf* function. Its purpose is to verify that the functions used to calculate the Jacobians are accurate (Section 3.5.3). To do this, YAO relies on the functional derivative (called Fréchet derivative (Fréchet, 1941)) and the directional derivative (called Gateaux derivative (Gateaux, 1919)). If the differences between the values obtained by YAO and the test do not exceed the precision provided as input, the test is considered conclusive. The second test (*testof*) checks the derivatives of the cost function (Section 3.5.2). There is also the adjoint test (*testad*) and the tangent linear test (*testlt*), (Section 3.5.1) acting to verify the accuracy of the transposition of the tangent linear and to verify the accurateness of the tangent linear model.

4.4. Development of SECHIBA-YAO 1D

The version of SECHIBA implemented in YAO includes the two-layer hydrology presented in Section 2.2.3. SECHIBA original code is implemented in a modular scheme, having a set of well-defined routines, independent in its processes and with a single entry point (a main routines handling the rest of the functionalities). The version of SECHIBA chosen works in a one dimension.

A set of prognostic variables is defined for each module and its assignation depends on the forcing conditions, physics phenomena, etc. SECHIBA can work coupled with the other components of ORCHIDEE (STOMATE and LPJ) or it can be used offline, as it was used in this work. Once SECHIBA is coded in YAO, it can be easily coupled with the other modules of ORCHIDEE.

In SECHIBA, the different routines were coded using Fortran language and it runs at any resolution and over any region of the globe. As follows, the version of SECHIBA implemented in YAO is called SECHIBA-YAO 1D and the original version of the model, coded in Fortran, is called SECHIBA-Fortran.

ORCHIDEE uses MODIPSL and IOIPSL in its internal processes (see <http://forge.ipsl.jussieu.fr/igcmg/wiki/platform/documentation> for more information). Developed in IPSL, the first one is a set of scripts allowing the extraction of a given configuration from a computing machine and the compilation of the specific machine configuration components. MODIPSL is the tree that will host models and tools for configuration. IOIPSL helps to manage variables state history, variable normalization, file lecture, and among others.

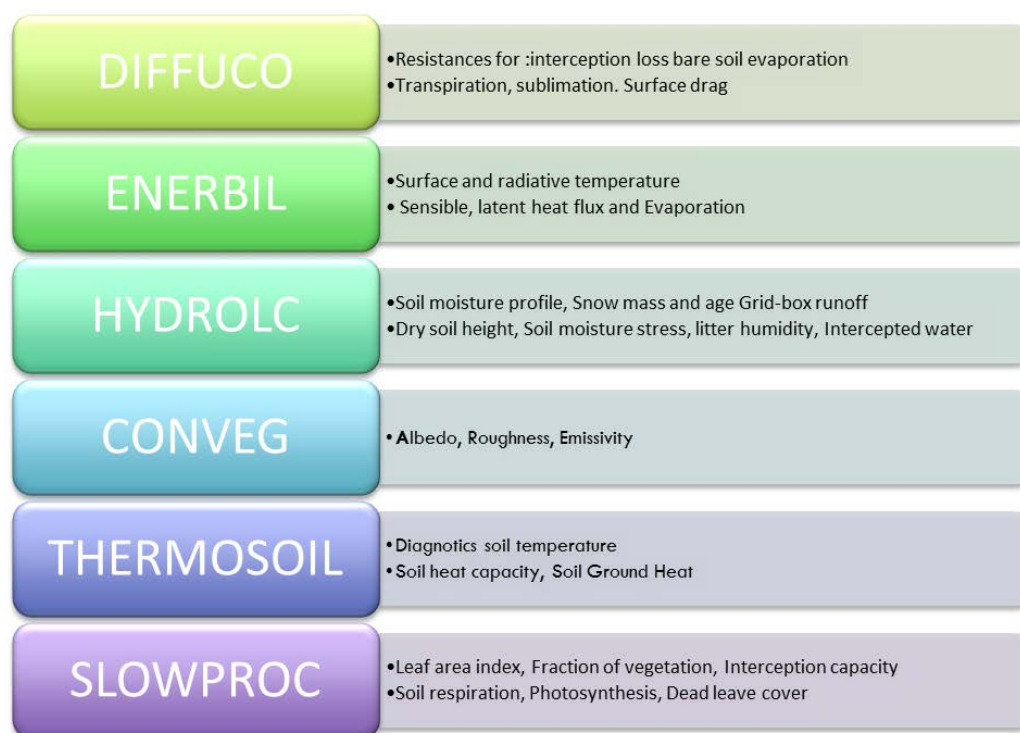


Figure 4. 3 SECHIBA subroutines and its corresponding outputs

The main routines in SECHIBA-Fortran are presented in Fig.4.3. These are the routines considered in the YAO implementation of the model. First, DIFFUCO computes the diffusion and plant transpiration coefficients based on the atmospheric conditions, solar fluxes, dry soil height, soil moisture stress and fraction of vegetation. ENERBIL corresponds to the energy budget module. Surface energy fluxes related to the soil are computed, based on atmospheric conditions, radiative fluxes, resistances, surface type fractions and surface drag. HYDROL

is the hydrological budget module, taking as inputs the rainfall, snowfall, evaporation components, soil temperature profile and vegetation distribution. CONDVEG helps in the computation of the vegetation conditions. The thermodynamics of the model is computed in THERMOSOIL, based on a seven-layer soil profile. Finally, SLOWPROC calculates the soil slow processes. When SECHIBA is decoupled from STOMATE, this module deals also with the LAI evolution.

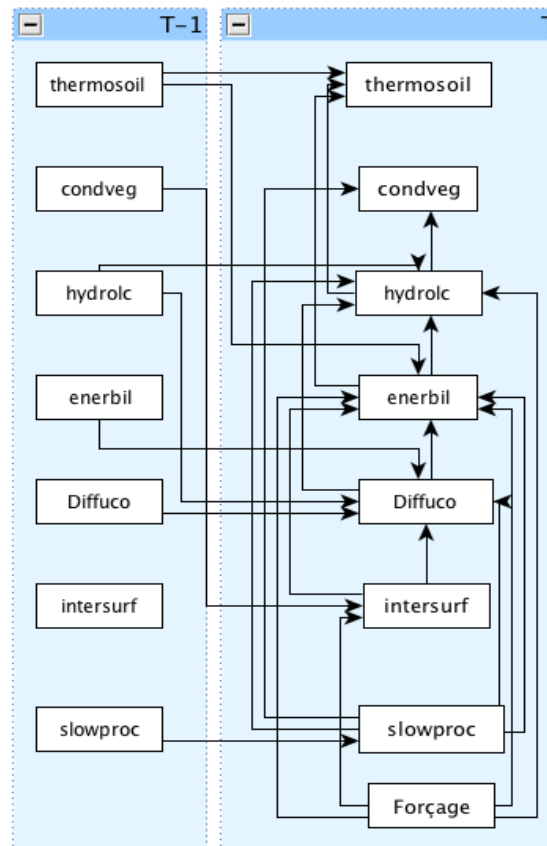


Figure 4. 4 SECHIBA hyper graph, showing general model dynamics

The different SECHIBA components are interconnected as shown in Fig.4.4. The output of the different modules serves as inputs for the next one, thus resulting in an interdependency among modules to be considered when modeling SECHIBA-YAO 1D. A more detailed graph representation of every main routine can be found in Appendix B.

4.4.1 Implementation Outline

Modular Graph

The implementation of SECHIBA in YAO started with the definition of the modular

graph describing the dynamics of the model. Elementary processes and interconnections between modules were defined in order to catch the essence of the model. This is a crucial step since the modular graph is the basis of all the integration processes made by YAO. Direct and adjoint model are computed following the modular graph structure. Consequently it is imperative to define the right dynamics of the model.

With SECHIBA, the modular graph was built as follows (this list represents the thesis work):

1. Every component of the original code was studied in detail line by line directly looking at the original FORTRAN code. The code corresponding to the IPSL libraries (MODIPSL and IOIPSL) was ignored. The remaining 25000 lines of code corresponding to SECHIBA were processed.
2. The main routine in SECHIBA that calls the other subroutines (Fig.4.3) states the model dynamics; hence, the order the subroutines are called was obtained (Fig.4.4).
3. A list of inputs and outputs for each subroutine was made, for every routine mentioned in Fig.4.3. With this, information flows in the model is exactly known.
4. A second zoom in the subroutines was made in order to understand the internal dynamics. This is the last step in the modular graph definition. When studying the subroutines, they were very general and a division into simpler elements was inevitable, with the purpose of reducing the coupling and increasing the cohesion of the modules. The idea is to have a scalable code. Uncoupled modules give more independence when changing part of the model. Cohesive modules help to understand the model.
5. From the initial six subroutines in SECHIBA-Fortran, SECHIBA-YAO 1D modular graph has 130 modules, corresponding to every process modeled by SECHIBA and a number of temporal modules serving as auxiliary computing.
6. It is important to mention that every variable and subroutine name was kept as in the original model. If a user or developer of SECHIBA-Fortran sees the implementation in YAO, he will find his way easily.

Direct model

After defining the modular graph in YAO, the second step in the implementation outline is the coding of the direct and the derivatives of the modules. The user has two options: a mapping between modules and external executables, when they must have the same number

of inputs and outputs as the module declared in YAO. The second consist in coding the different modules directly with YAO meta-language. Every module will be represented as a script and the different processes attributed to the module are implemented inside the script, allowing a better control of the physics, i.e. any change in the physics could be made without the initial FORTRAN code. In SECHIBA-YAO 1D, the second approach was used.

For each subroutine, a set of modules was defined, as a decomposition of the main subroutine (see Appendix B). Table 4.1 shows the partition made for each subroutine.

Subroutine	Number of modules
DIFFUCO	15
ENERBIL	22
HYDROLC	32
CONDVEG	2
THERMOSOIL	7
SLOWPROC	5

Table 4. 1 SECHIBA Subroutines and number of modules

Given the complexity of some subroutines, particularly ENERBIL and HYDROLC, the partition was more important than other subroutines. In effect, both subroutines have a long list of processes and its outputs are the core base of SECHIBA. The other subroutines only compute ancillary variables to calculate the energy and hydrologic budget.

Hereafter, for every module, its process was coded line-by-line, migrating the original Fortran code. This coding process introduced some errors: conditions coded inaccurately, transposition of inputs and outputs in the module or in the modular graph description file, misinterpretation of the original code, portions of code ignored, among others; thus, an important debugging work was performed. The validation of the direct model results are presented in the section 4.4.2.

Module Derivatives

Once the direct model was coded and validated, the next step in the implementation process is the derivative calculation of every module defined in SECHIBA-YAO 1D. There are two options to code the derivatives: they can be coded line-by-line based on the forward computing, in order to obtain the Jacobian matrix of the module, or they can also be produced routinely , using an automatic differentiation tool (for example, Tapenade (Hascoët et al, 2012)).

For this work the first option was applied given that the processes in the different modules were not difficult to differentiate, the developer preferred to code by itself the Jacobian matrix of the modules. The derivative process is made line-by-line. The outputs are derived with respect to every input. YAO will generate automatically, based on these derivatives, the tangent linear and the adjoint of the model.

Nevertheless, the derivative process introduced errors related to the coding process. Besides the same kind of errors mention it in the previous point, errors related to the derivative process were also introduced: inexact derivatives, expressions that were not differentiated among others. In order to reduce it to a minimum number of bugs, the adjoint of the model was validated (as it was made with the direct model). This guarantees the accuracy when performing assimilation. The validation of the adjoint model is presented in Section 4.4.3 and in Chapter 5 (Sensitivity Analysis), where several techniques are presented with the aim of validating the adjoint model.

Once the derivatives of the model are validated, the adjoint model made by YAO is correct and it can be used to perform VDA. The next step is the definition of assimilation scenarios in order to study the potential of the model. To do so, several assimilation scenarios are defined in order to establish different experiments. These experiments are presented in the Chapter 6 Twin Experiments, where a complete set of experiments are presented.

4.4.2 Direct model Validation

We had to make sure that SECHIBA-YAO 1D is equivalent to SECHIBA-Fortran. At the end of the migration process, a validation with respect to SECHIBA-Fortran was a requirement.

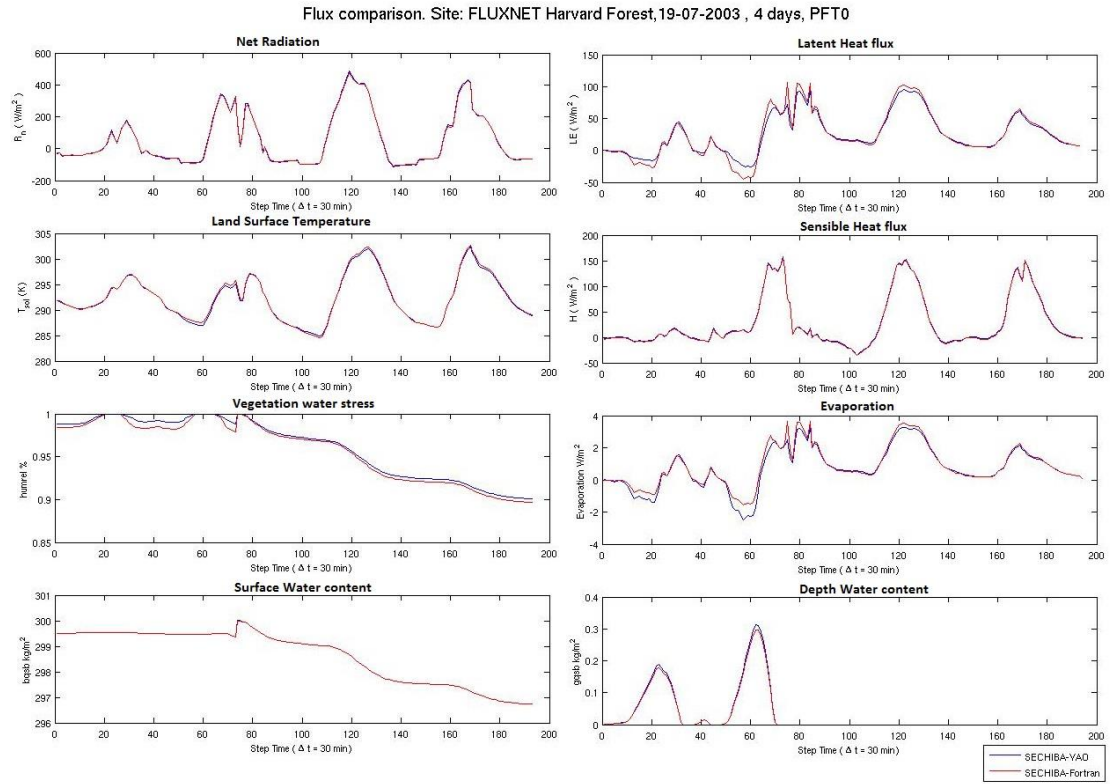
Diagnostic variables	Kruger Park				Harvard Forest			
	PFT 0		PFT 11		PFT 0		PFT 11	
	RMSE	Relative error	RMSE	Relative error	RMSE	Relative error	RMSE	Relative error
Net Radiation	0,0190	0,0013	0,0729	0,0558	0,0215	0,012	0,0315	0,023
Sensible heat flux	0,0509	0,035	0,0795	0,0658	0,075	0,062	0,071	0,0501
Latent heat Flux	0,0711	0,0507	0,0841	0,0715	0,0718	0,055	0,0765	0,058
Soil Evaporation	7,51e ⁻⁴	1,8e ⁻⁵	0,0042	0,0035	2,6e ⁻⁴	1,89e ⁻⁴	4,2e ⁻⁴	2,5e ⁻⁴
Transpiration	-	-	0,0025	0,0015	-	-	7,1e ⁻⁴	5,1e ⁻⁴
Land surface temperature	6,61e ⁻⁴	5,74e ⁻⁴	8,7e ⁻⁴	5,01e ⁻⁴	6,34e ⁻⁴	5,17e ⁻⁴	1,28e ⁻³	7,01e ⁻⁴
Soil Moisture	3,62e ⁻⁴	9,33e ⁻⁴	2,5e ⁻⁴	1,5e ⁻⁴	2,1e ⁻⁴	1,25 e ⁻⁴	6,1e ⁻⁴	5,9e ⁻⁴
Soil Water Content (surface layer)	0,0033	0,0034	0,0025	0,0015	0,0070	0,0065	0,006	0,0051
Soil Water Content (ground layer)	0,0070	0,0012	0,0021	0,0014	0,0039	0,0025	0,0069	0,0062

Table 4. 2 RMSE and Relative Error between SECHIBA-Fortran and SECHIBA-YAO 1D. These errors were computed based on a year simulation on both models.

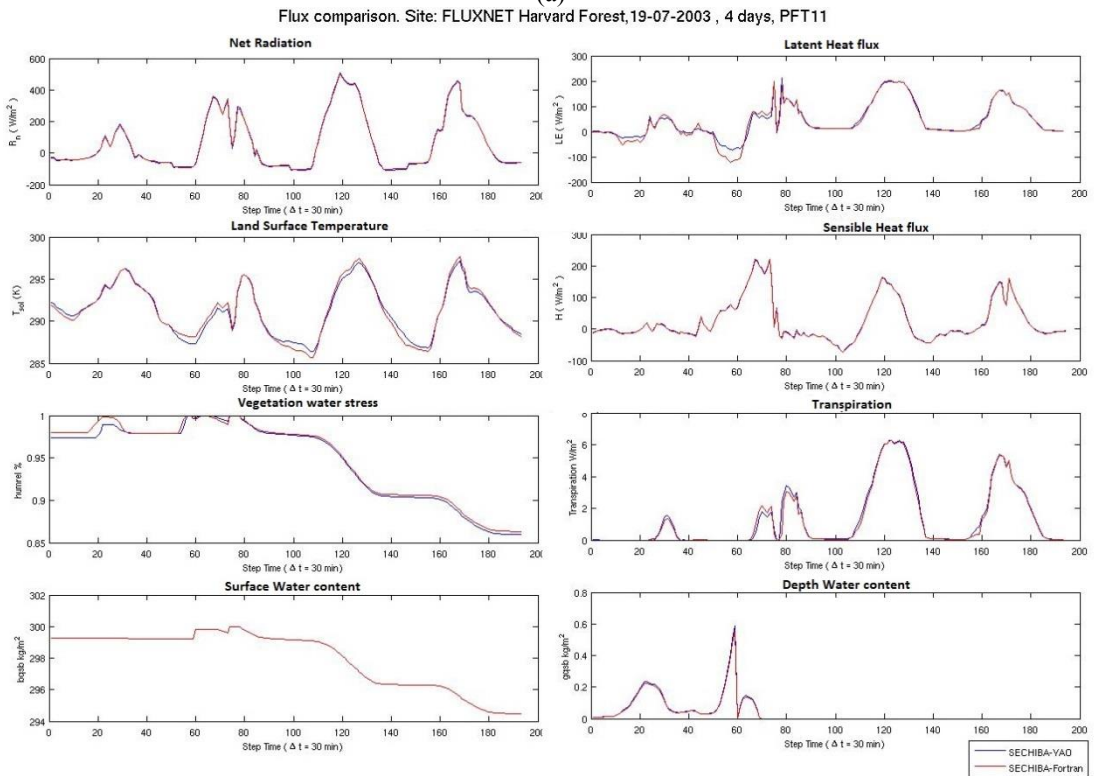
Both models were run one year for two different scenarios, using the FLUXNET Harvard Forest 1996 and Kruger Park 2003 forcing. For nine diagnostic variables, the root mean square error and the relative error were computed, as presented in Table 4.2, for one type of vegetation (grassland PFT 11) and bare soil (PFT 0).

Even though small differences are found, the simulations made by SECHIBA-YAO are close enough to SECHIBA-Fortran to conclude that the direct model is valid. Both models predict the same flux behavior under several forcing scenarios, even though those small disparities come up with some extreme cases, the general performance is kept in SECHIBA-YAO 1D.

In addition, Fig.4.5 and Fig.4.6 show the time series of several variables during four days, using Harvard Forest and Kruger Park forcing, respectively. For the different scenarios presented, eight diagnostic variables were compared: net radiation (R_n), latent heat flux (LE); land surface temperature (LST), sensible heat flux (H), vegetation water stress, transpiration, surface water content and depth water content. It can be seen that the differences are almost indiscernible for several of them. Although small differences between both results are more perceptible for several curves in this four day time series, in a scale of a year, the accuracy of SECHIBA-YAO 1D do not vary too much (Table 4.2).



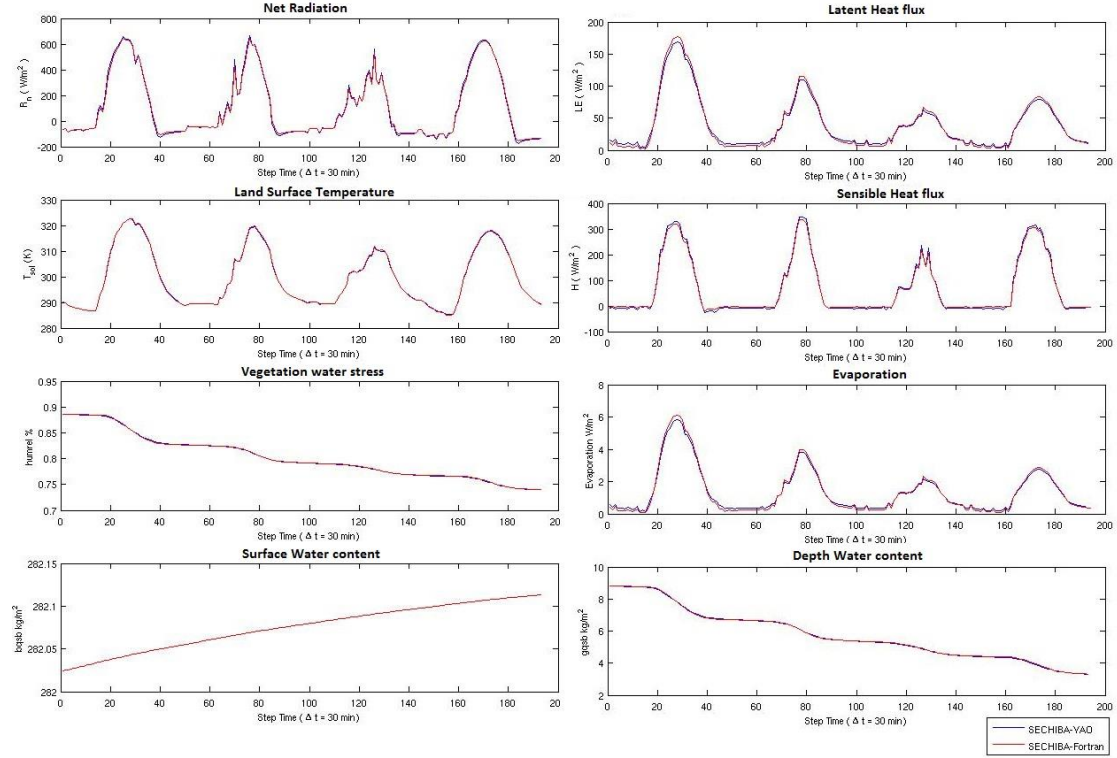
(a)



(b)

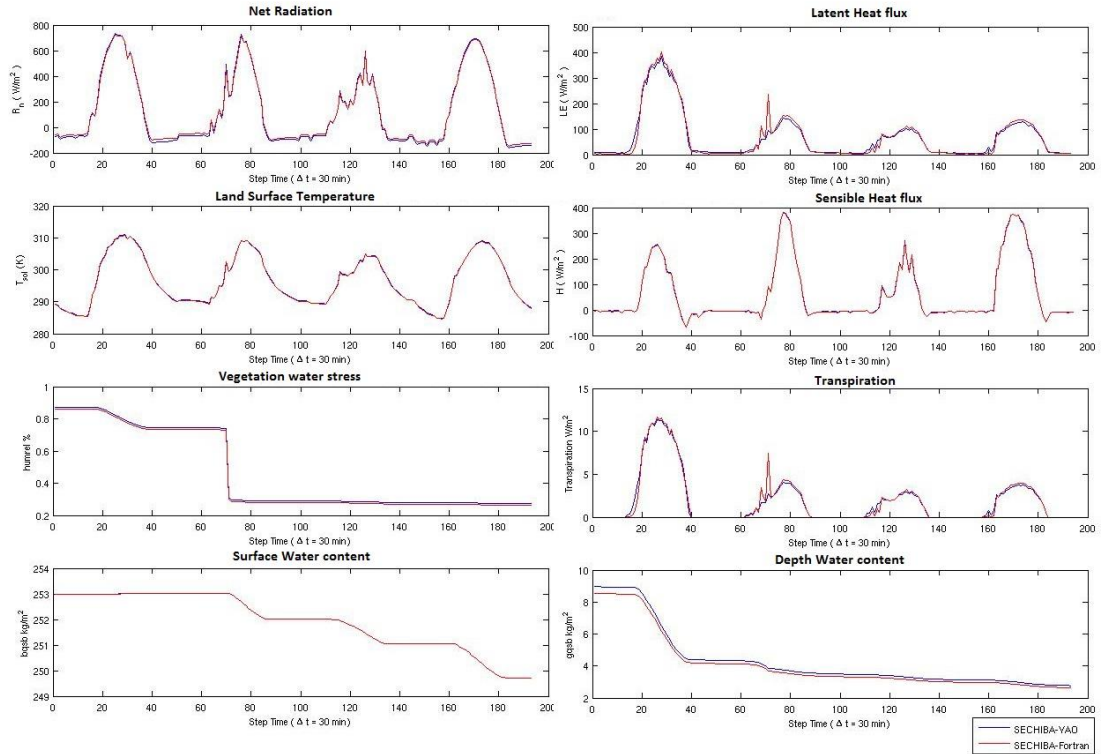
Figure 4.5 Comparison between SECHIBA-Fortran and SECHIBA-YAO 1D outputs, using FLUXNET Harvard Forest forcing from 19/07/2003 to 23/07/2003. Each curve is a time series during 4 days for 8 diagnostic variables. Red curves are the estimations computed with SECHIBA-Fortran. Blue curves represents SECHIBA-YAO -1D estimation. (a) is with bare soil and (b) with grassland.

Flux comparison. Site: FLUXNET Kruger Park, 7-02-2003 , 4 days, PFT0



(a)

Flux comparison. Site: FLUXNET Kruger Park, 7-02-2003 , 4 days, PFT11



(b)

Figure 4.6 Comparison between SECHIBA-Fortran and SECHIBA-YAO 1D outputs, using FLUXNET Kruger Park forcing from 07/02/2003 to 11/02/2003. Each curve is a time series during 4 days of 8 diagnostic variables. Red curves are the estimations computed with SECHIBA-Fortran. Blue curves represents SECHIBA-YAO -1D estimation. (a) is with bare soil and (b) with grassland.

4.4.3 Adjoint Model Validation

Once the direct model is validated and the derivatives of each module are computed, YAO generates (while compiling the project) a code that represents the backward integration of the modular graph. In order to check that these derivatives are correctly computed, the adjoint model was verified in two phases: the first consists in the implementation of the tests mentioned in Section 3.4, where different diagnostic tools allow us to determine if the adjoint model is accurate. In order to do so, these tests are already implemented in YAO, thus the results for the adjoint test, cost function test and derivative tests are presented in the next section. The second phase, which is presented in Chapter 5, Sensitivity analysis, consists in a gradient sensitivity study concerning the land surface temperature with respect to the control parameters. Local sensitivity analysis was computed on SECHIBA-YAO 1D using the adjoint model and the results were compared to the gradients calculated using the direct model (finite differences) in order to test the exactitude of SECHIBA-YAO 1D adjoint to deliver adequate values

Derivative test

The derivative test, as it was mentioned in Section 3.4.3, consist in verifying for each input of each module that the derivative coded in the module is equal to an approximation computed in YAO, (Eq.3.13). The results are not presented since there are very extensive. However, all modules passed the test, and thus it can be said that the different derivatives coded in each module are valid.

Adjoint test

The adjoint test, as it was mentioned in Section 3.4.1 indicates the consistency between the linear tangent model (*forward*) and the adjoint model (*backward*), by verifying the scalar product mentioned in Eq.3.7. The results of applying this test to SECHIBA-YAO 1D are presented in the Fig.4.7. It can be observed that both terms are equal up to 16 digits, thus the test is verified.

$\langle \mathbf{M}^T(x_0).dx, dy \rangle$	9.150614798923334e-01
$\langle dx, \mathbf{M}^T(x_0).dy \rangle$	9.150614798923330e-01
Difference	4.440892098500626e-16
Relative error	4.853108e-16

Figure 4. 7 Results of the adjoint test applied to SECHIBA-YAO 1D. The difference between both terms is presented in the lower part of the figure.

Cost Function test

The third test presented in this section concerns the calculation of the cost function gradients, based on the test presented in Section 3.4.2. As it was mentioned before, the idea of this test is to compute the gradients of the cost function in successive iterations, by reducing a factor α . The gradient must point in the same direction as the difference between two realizations of the cost function which are separated by a small perturbation in model state.

The test was applied using Kruger Park forcing with grassland (PFT 11), from 07/02/2003 to 11/02/2003. A perturbation of 1% was applied equally to the different control parameters. The α initial value is 1 and it is reduce by a factor of 10 every iteration. The results of this test given by YAO are presented in Table 4.3.

Cost function: $J(x_0) = 3.745172e-01$			
It #	α	Order 1 $\rightarrow 1$	Order 2 $\rightarrow K$
1	1.000000e-01	7.352809e+00	7.342809e+03
2	1.000000e-02	6.729459e+00	7.342459e+03
3	1.000000e-03	8.306182e+00	7.326182e+03
4	1.000000e-04	8.368875e+00	7.318752e+03
5	1.000000e-05	9.760466e-01	7.139534e+03
6	1.000000e-06	9.170205e-01	7.029795e+03
7	1.000000e-07	9.077703e-01	2.922297e+05
8	1.000000e-08	9.208394e-01	2.791606e+06
9	1.000000e-09	9.017666e-01	2.982334e+07
10	1.000000e-10	9.017855e-01	2.982145e+08

Table 4. 3 Cost function test results given by YAO.

In Table 4.3, results of applying the cost function gradient test are presented. The *order 1* column values must tends towards unity and order 2 column values must tend towards a constant K proportional to the Hessian matrix. It can be observed that the first four iteration values verified this condition. However, from the 6th iteration onward, the test diverges for the order 2: K values do not tend towards constant values. Since the calculation of the *order 2* criteria involves the algebra of very similar quantities, these operations can induce a loss of accuracy in the results. In order to pursue this idea, a numerical analysis validation was implemented in order to account for the significant digits obtained with the different

operations involving the cost function gradient test.

Numerical validation of SECHIBA-YAO 1D

Significant digits represent the number of decimal digits in common between a computer solution and the results corresponding to the exact mathematical result. To properly quantify the accuracy of a computed result, a tool called CADNA is used to retrieve the exact significant digits, for every computation made in SECHIBA-YAO 1D. CADNA, means Control of Accuracy and Debugging for Numerical Applications. It was developed in LIP6 laboratory in France. The main objective of CADNA is the accuracy estimation of a computed result. It uses concepts and definitions taken from stochastic arithmetic (Vignes, J., 1993) regarding order relations and equality relations. It includes the stochastic definitions of all the elementary arithmetic operations, order relations and elementary functions defined for the classical numerical types.

CADNA detects numerical instabilities in run time. This debugging is focused in the computer capability to give correct results when the code is running. The accuracy estimation is available for any intermediate or final result. The exact significant digits of every operation are displayed as output. CADNA is able to detect when the conditions for a right estimation of the round-off errors are not satisfied anymore and when it happens, CADNA is able to advise the users. Therefore, the numerical debugging and the self-validation method are performed by systematically detecting numerical instabilities.

Gradient Test

The gradient test helps to verify the accuracy of the model gradient by computing the residue. The idea is to calculate the variation of the residue defined by Eq.4.1, depending on the amplitude α of the perturbation imposed on the control variable. If the gradient is valid, the residue should vary as α^2 , meaning as the calculation makes sense on the machine, that is to say, as long as the residue is greater than the product $\varepsilon.J(x_0)$ or as long as the residue has still some significant digits, where ε is the “machine zero” ($\varepsilon = 2^{-53}$). The residue must be quadratic, which proves that the gradient is correct (strictly speaking in the direction of disturbance dx). The residue is computed with the following expression

$$R(\alpha) = [J(x_0 + \alpha dx) - J(x_0)] - \alpha \nabla J(x_0) dx \quad (4.1)$$

In Eq.4.1, J is the cost function, ∇J is the gradient of the cost function with respect to x , dx is the perturbation vector and α is the amplitude of the perturbation. The inputs

necessary in order to perform this test are taken from the cost function results obtained, presented in Table 4.3. This test was applied using Kruger Park forcing with grassland (PFT 11), for a different date from the one used in Table 4.3, (from 02/02/2003 to 07/02/2003). A perturbation of 1% was applied equally to the different control parameters. This test was performed with and without CADNA, in order to assess the numerical precision lost when computing the residue.

In Fig. 4.7, the gradient test results are shown. The black dotted line represents α^2 , red dotted line is α and the blue line is the residue, for different values of the amplitude α , going from 10^{-1} to 10^{-16} . Logarithmic scales were used throughout. As it can be seen in Fig. 4.7, the first six values of α are quadratic (blue line varies as α^2), however from the 6th iteration onwards, the gradient diverges. From the moment the test diverges in Fig. 4.7.(b), CADNA finds zero significant digits (Fig. 4.7.(a)), showing that there is no precision left from that point in the residue calculation. Looking closer at the precision of the residue equation, two different subtractions introduce a precision lost in the result.

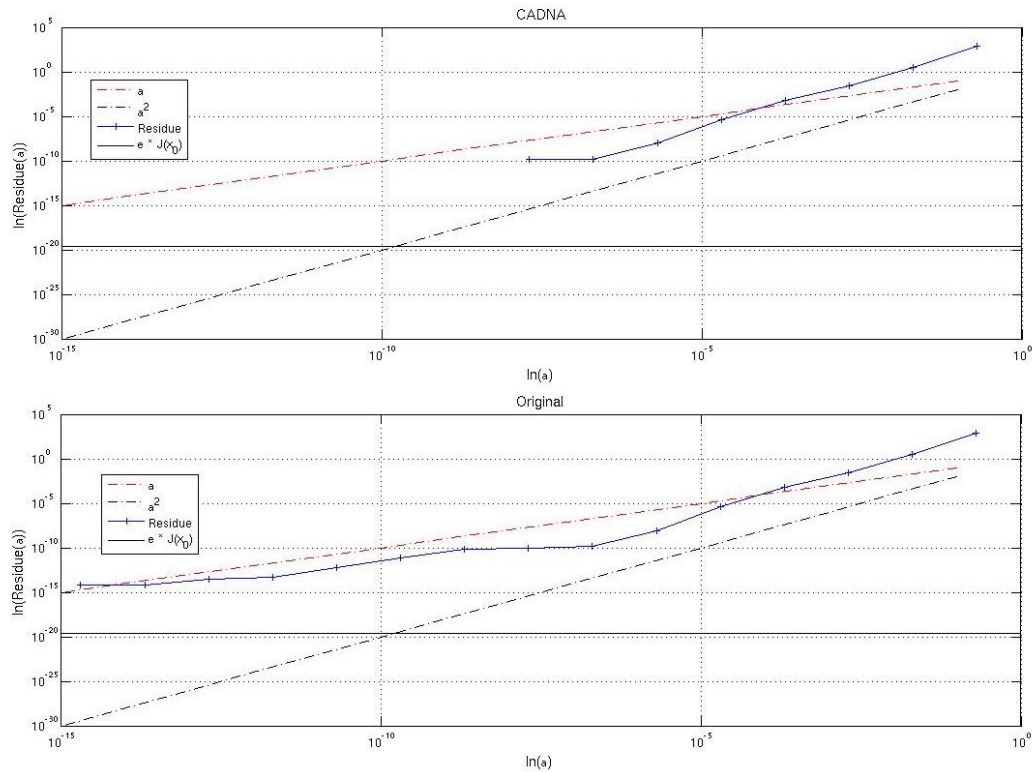


Figure 4. 8 Gradient test result using Kruger Park forcing from 07/02/2003 to 11/02/2003, with and without CADNA with grassland (PFT 11). Residue curve (blue) must vary as α^2 (black dotted line) so that the test is valid.

Even though the results without CADNA clearly show a divergence in the residue

(Fig.4.7.(b)), the results obtained with CADNA are trust worthier. These clearly indicate a loss of accuracy, meaning that these results cannot be taken into account in our analysis of the gradient. On the contrary, within the first six iterations, CADNA found significant digits indicating that the gradient test is valid, since these increase as α^2 .

The cardinal rule of numerical computing aims to avoid subtracting nearly equal numbers. The closer the two numbers are, the more precision is lost in the subtraction. For example, if we have two values x and y and they agree to m bits, up to m bits of precision can be lost in computing $x-y$. This behavior can be found in the residue formula: when α becomes smaller, the two members of the equation $\alpha \nabla J(x_0).dx$ and $[J(x_0 + \alpha.dx) - J(x_0)]$ have almost the same value. From the 6th iteration onwards, their value agrees to 12 decimal places, so about 12 decimal places are lost in the subtraction. A double precision floating point number (as it was declared for all of our variables) contains about 15 significant digits (corresponding to machine epsilon). So when increasing α , the rounding error will be higher than 10^{-15} and thus we lost all the precision when computing the residue.

Chapter 5

Sensitivity Analysis of the SECHIBA-YAO 1D model using FLUXNET dataset

5.1 Introduction

Sensitivity analysis is the study of the variation of the output of a model (numerical or otherwise) with respect to the input (the forcing) or to the parameterization of the different processes represented by the model (Saltelli et al 2008). It is not concerned with what causes the output of the model to be what it is, but what the sources of variation are. In the present study, sensitivity refers to the local impact of the parameters on the model, providing the gradients of the calculated model output in the parameter space for a given set of values. It also helps to analyze the behavior of the system for a trajectory in the phase space defined by the prior values of the parameters. Local analysis has the advantage of facilitating the mechanistic interpretation of sensitivities (sign and magnitude) (Saltelli et al., 2008).

The adjoint state method allows computing the scalar function derivatives for a calculation cost independent of the area dimension controlled (Castaing 2007). In the linear framework, as in the non-linear, interpretation of the adjoint variables permits an objective understanding of the relationship $k \rightarrow x \rightarrow J(x, k)$, in the form of

$$\frac{\partial J}{\partial k} = \frac{\partial J}{\partial x} \cdot \frac{\partial x}{\partial k} \quad (5.1)$$

with k the parameter vector, x the state of the model and J a cost function. In a temporal framework, the variational approach is used to calculate the effect on the control variables of

time dependent disturbance of the input.

Several works regarding the implementation of local sensitivity analysis prior to assimilation conclude that sensitivity analysis is a mandatory step in the model building process. Greenwald et al. 2004 worked with a regional atmospheric modeling system in order to explore the connection between cloud microphysics and top of the atmosphere radiances. These authors developed an adjoint sensitivity analysis scheme for an observation operator in order to understand the potential of the different measurements available and the influence the parameters have in the variable estimation. They conclude that sensitivity analysis allowed identifying a quantifiable relationship between model parameters and observation operators. In addition, Breierova and Choudhari, 1996 mentioned that sensitivity analysis helps the modeler to understand the dynamics of a system. Experimenting with a wide range of values can offer insights into the behavior of a system in extreme situations. Discovering that the system behavior varies for a change in parameter values can identify leverage points in the model. In addition, the authors in this work stated that sensitivity analysis helps to build confidence in the model by studying the uncertainties that are often associated with model parameters.

Prior to a calibration exercise, sensitivity analysis can be employed, as explained in Giering et al. 1998, to investigate the importance of tuning each parameter in the model, *i.e.* to identify a candidate set of important factors for calibration, since the difficulty of calibrating models against field or laboratory data increases with the number of processes to be modeled (and hence the number of parameters to be estimated).

When building a numerical model, modelers can conduct sensitivity analysis (Saltelli et al 2008) to determine:

1. If a model accurately reproduces the system or the processes under study.
2. The factors (parameters or initial conditions) that mostly contribute to the output variability.
3. The model parameters that are insignificant, and that can be eliminated to simplify the final model.
4. If there is some region in the space of parameters for which the model sensitivity is maximum and how the sensitivity depends on initial and boundary conditions.
5. The optimal regions within the space of the factors to be used in a subsequent calibration

study.

6. Parameters interacting with each other.

In this work, our efforts are focused on points 2, 3, and 4. Once SECHIBA-YAO 1D implementation is achieved and the direct model validated (Chapter 4), a sensitivity analysis with two main objectives is performed which is reported in this Chapter: to test the accuracy of the adjoint method and to determine a parameter hierarchy of the most sensitive parameters to the estimation of land surface temperature.

5.2 Variational Sensitivity Analysis

The sensitivity analysis we did was based on the use of the adjoint model of SECHIBA 1D. It has been the subject of an article submitted to the Journal of Geophysical Research (JGR). This article is included in the following section.

5.2.1. Sensitivity analysis with land surface temperature

Article Summary

In this paper, a variational sensitivity analysis was done, prior to the implementation of variational data assimilation, by using the adjoint model of SECHIBA, generated with the adjoint semi-generator software YAO. Once the parameter hierarchy is obtained with the sensitivity analysis, twin experiments using synthetic observations are presented, allowing us to evaluate the model response to the assimilation process. The results obtained when controlling the most sensitive parameters and the initial soil water content show the flexibility of the assimilation scheme.

Article

Variational Assimilation of Land Surface Temperature within the ORCHIDEE Continental Surface model

H. S. Benavides Pinjosovsky^{1,2,3}, C. Ottlé², S. Thiria¹, P. Maugis², J. Brajard¹, and F. Bradran¹

[1]{Laboratoire d'Océanographie et du Climat: Expérimentations et approches numériques, IPSL Paris, France}

[2]{Laboratoire des Sciences du Climat et de l'Environnement, IPSL, CNRS-CEA-UVSQ, Gif-sur-Yvette, France}

[3]{CLIMMOD Engineering, Orsay, France}

Correspondence to: H. S. Benavides Pinjosovsky (hbplod@locean-ipsl.upmc.fr) and C. Ottlé (catherine.ottle@lsce.ipsl.fr)

Abstract

Variational data assimilation is applied to the energy and water budget modules of the ORCHIDEE land surface model. This part of the model called SECHIBA, describes the exchanges of water and energy between the surface and the atmosphere. The adjoint semi-generator software called YAO was used as a framework to implement 4D-VAR assimilation. First, sensitivity analysis was performed in order to validate the adjoint model and to identify the most influential parameters. Then, the results of twin experiments using synthetic observations are presented in order to demonstrate the robustness of the assimilation. The results obtained when controlling the most sensitive parameters and the initial soil water content show the flexibility of the assimilation scheme and the potential of land surface temperature variational data assimilation to improve model calibration and reduce prediction errors.

Keywords: Sensitivity Analysis, Data Assimilation, Model Calibration, Land Surface Temperature

1. Introduction

Land surface models (LSM) simulate the interactions between the atmosphere and the land surface, which influence directly the exchange of water, energy and carbon with the atmosphere. They are important tools for understanding the main interaction and feedback processes simulating the present climate and making predictions of future climate evolution (Harrison et al., 2009). Such predictions are subject to considerable uncertainties, related to the difficulty to model the highly complex physics with a limited set of equations that does not account for all the interacting processes (Pipunic et al., 2008, Ghent et al. 2011). Understanding these uncertainties is important in order to obtain more realistic simulations.

The main challenge of each dynamical model, regardless its nature, is to have the appropriate source of information to produce an accurate response. Observations sample the system of interest in space and time. These measurements provide essential information on the model dynamics and contribute to the understanding of the system evolution (Lahoz et al. 2010). Data assimilation adds observations to the model, allowing extracting valuable information. The idea is to merge the measurements with the dynamical model with the purpose to arise a more accurate estimate of the current and future states of the system, together with uncertainty estimates in the model states. Two basic methodologies can be used to come up with uncertainties. The sequential approach (Evensen 2003), based on the statistical estimation theory and the Kalman filter, and the variational approach (Le Dimet et al., 1986), (4DVAR) built from the optimal control theory (Robert et al, 2007). It is well known that both approaches provide the same solution at the end of the assimilation period, for perfect and linear models. These approaches are different, however, mainly because the model is seen as a strong constraint in the 4DVAR approach and as a weak constraint in the sequential approach. Variational data assimilation has been widely used in land surface applications. The assimilation of land surface temperature

(LST) is suitable for an extensive range of geophysical problems. As mentioned in Ridler et al. (2012), LST is an excellent candidate for model optimization since it is solution of the coupled energy and water budgets, and allows us to constrain parameters related to evapotranspiration and indirectly to soil water content. In Castelli et al. (1999), a variational data assimilation approach is used to include surface energy balance in the estimation procedure as a physical constraint (based on adjoint techniques). The authors work with satellite data, where soil skin temperature is directly assimilated. They conclude that constraining the model with such observation improves model flux estimates, with respect to available measurements. In Huang et al. (2003) the authors developed a one-dimensional land data assimilation scheme based on ensemble Kalman filter, used to improve the estimation of land surface temperature profile. They demonstrate that the assimilation of LST into land surface models is a practical and effective way to improve the estimation of land surface state variables and fluxes. Reichle et al. (2010) performs the assimilation of satellite-derived skin temperature observations using an ensemble-based, offline land data assimilation system. Results suggest that retrieved fluxes provide modest yet statistically significant improvements. However, these authors noted strong biases between LST estimates from *in situ* observations, land modeling, and satellite retrievals that vary with season and time of day. They highlighted the importance of taking these biases into account. Otherwise large errors in surface flux estimates can result. Ghent et al. (2011) investigated the impacts of data assimilation on terrestrial feedbacks of the climate system. Assimilation of LST helped to constrain simulations of soil moisture and surface heat fluxes. Ridler et al. (2012), tested the effectiveness of using satellite estimates of radiometric surface temperatures and surface soil moisture to calibrate a Soil–Vegetation–Atmosphere Transfer (SVAT) model, based on error minimization of temperature and soil moisture model outputs. Flux simulations were improved when the model is calibrated against *in situ* surface temperature and surface soil moisture versus satellite estimates of the same fluxes. In Bateni et al. (2013), the full heat diffusion equation is employed in the variational data assimilation scheme as an adjoint (constraint). Deviations terms of the evaporation fraction and a scale coefficient are added as penalization terms in the cost function. Weak constraint is applied to data assimilation with model uncertainty, accounting in this way for model error. The cost function to this experiment contains a term that penalizes deviation from prior values. When assimilating LST into the model, the authors proved that the heat diffusion coefficients are strongly sensitive to specific deep land surface temperature. As a conclusion, it can be seen that the assimilation of LST can achieve a remarkable improvement in the model simulated flows. Since many studies demonstrate the usefulness of LST data assimilation, this paper explores the potential of 4DVAR data assimilation in constraining ORCHIDEE model.

In the present study, a variational data assimilation scheme was implemented in the ORCHIDEE Land Surface Model, developed at the “Institut Pierre Simon Laplace (IPSL)” in France. In the early stages of this work, efforts were concentrated on the biophysical model denoted SECHIBA (Ducoudré et al. 1993), which is part of ORCHIDEE. SECHIBA describes the exchanges of water and energy between the surface and the atmosphere. The assimilation was conducted by using the adjoint semi-generator software called YAO developed at LOCEAN-IPSL (Nardi et al. 2009). YAO serves as a framework to design and implement dynamic models, helping to generate the adjoint of the model allowing computing the model gradients. Sensitivity analysis was performed in order to test the adjoint of SECHIBA and to identify the most influential parameters in the LST estimation. This is a crucial step prior to assimilation (Barrett et al. 2009, Ridler et al. 2012).

Model parameters as well as initial conditions of surface soil water content were controlled in our

assimilation scheme. Sensitivity analysis were performed for different soil and climate conditions, i.e., for a bare soil and a grassland, forced with two different climates provided by two FLUXNET meteorological stations (Baldocchi et al, 2001). Next, twin experiments were performed in order to test the robustness of the assimilation with our model and to assess the potential of LST assimilation. In addition, the impact of increasing the number of control parameter set was tested. The final objective of these experiments is to analyze the potential of implementing 4DVAR into SECHIBA, in order to quantify the robustness of the assimilation process to improve the estimation of model parameters and reduce model fluxes error, and the capability of LST in constraining SECHIBA parameters.

2. Models and Data

2.1. ORCHIDEE

ORCHIDEE is a mechanistic dynamic global vegetation model (Krinner et al., 2005) representing the continental biosphere and its different processes. It is part of the IPSL (Institut Pierre Simon Laplace) Earth system model (LMDZ, Hourdin et al., 2006; Dufresne et al., 2013) and is composed of 3 modules: SECHIBA, STOMATE and LPJ. SECHIBA computes the water and energy budgets at the biosphere-atmosphere interface, as well as the Gross Primary Production (GPP); STOMATE (Friedlingstein et al., 1999), is a biogeochemical model which represents the processes related to the carbon cycle, such as carbon dynamics, the allocation of photosynthesis respiration and growth maintenance, heterotrophic respiration and phenology and finally, LPJ (Sitch et al., 2003) models the global dynamics of the vegetation, interspecific competition for sunlight as well as fire occurrence. ORCHIDEE has different time scales: 30-minutes for energy and matter, one-day for carbon processes and 1-year for species competition processes. The full description of ORCHIDEE can be consulted in Ducoudré et al., 1993, Krinner et al., 2005, d'Orgeval et al., 2006, Kuppel et al., 2012. In the present study, ORCHIDEE 1.9 version is used in a grid-point mode (one given location), forced by the corresponding local half-hourly gap-filled meteorological measurements obtained at the flux towers. In this study, only the SECHIBA module is activated.

2.2 SECHIBA

SECHIBA (Schématisation des Echanges Hydriques à l'Interface Biosphère-Atmosphère) (Ducoudré et al., 1993) is a land surface model. It solves every half hour the energy budget of the surface and the soil water budget. The land surface is represented as a whole system composed of various fractions of vegetation types called PFT (Plant Functional Type). A single energy budget is performed for each grid point, but water budget is calculated for each PFT fraction. The resulting energy and water fluxes between atmosphere, ground and the retrieved temperature represent the canopy ensemble and the soil surface. The main fluxes modeled are the net radiation (R_n), soil heat flux (Q), sensible (H) and latent heat (LE) fluxes between the atmosphere and the biosphere, land surface temperature (LST) and the soil water reservoir contents. Energy balance is solved once, with a subdivision only for LE in bare soil evaporation, interception and transpiration for each type of vegetation.

Water balance is computed for each fraction of vegetation (Plant Functional Type or PFT) present in the grid. The SECHIBA version used in this work, models the hydrological budget based on a two-layer soil profile (Choisnel, 1977). The two soil layers represent respectively the surface and the total rooting zone. The soil is considered homogeneous with no sub-grid variability and a total depth of $h_{tot} = 2\text{m}$. The soil bottom layer acts like a bucket that fills with water from the top layer. The soil is filled from top to bottom with precipitation;

when evapotranspiration is higher than precipitation, water is removed from the upper reservoir. Runoff arises when the soil is saturated.

SECHIBA inputs are: R_{lw} the incoming infrared radiation; R_{sw} the incoming solar radiation; P the total precipitation (rain and snow); T_a the air temperature; Q_a the air humidity; P_s the atmospheric pressure at the surface and U the wind speed. No snowfall is taken into account in the present work.

2.3 FLUXNET data

FLUXNET (Baldocchi et al., 2001) is a network coordinating regional and global analysis of observations from micrometeorological tower sites. The flux tower sites use eddy covariance methods (Aubinet et al. 2012) to measure the exchange of carbon dioxide (CO_2), water vapor, and energy between terrestrial ecosystems and the atmosphere.

Measurement towers sprang up around the world, grouped in regional networks. The data from all networks is accessible to the scientific community via the Fluxnet website (<http://www.fluxdata.org>). In this work, we selected 2 sites: Harvard Forest and Skukuza Kruger National Park; both present contrasted climates and land surface properties to test the tools developed and assess model parameters sensitivities. Only climate measurements (model forcing) from both sites are used. Vegetation characteristics are prescribed and only homogeneous grids are considered. Two cases were studied with grassland (PFT 11) and bare soil (PFT 0). Forcing for both sites correspond to SECHIBA forcing, with the same sampling frequency (30 minutes).

Skukuza Kruger National Park

Located at latitude $25^\circ 1' 11''$ S and longitude $31^\circ 29' 48''$ E, in South Africa, this Fluxnet site was established in 2000. The tower overlaps two distinct savanna types and collects information about land-atmosphere interactions. The climate is Subtropical-Mediterranean. The total mean annual precipitation is 650 mm, with an altitude of 150 m and the mean annual temperature is 22.15°C .

Harvard Forest

Located in the United States of America, on land owned by Harvard University, the station is located at latitude $42^\circ 53' 78''$ N and longitude $72^\circ 17' 15''$ W. It was established in 1991. The site has a Temperate-Continental climate with hot or warm summers and cold winters. The annual mean precipitation is 1071 mm, the mean annual temperature is 6.62°C and the altitude is 340 m.

3. The Methodology

3.1 Variational assimilation

Variational assimilation (4D-VAR) (Le Dimet et al. 1986) consider a physical phenomenon described in space and its time evolution. It thus requires the knowledge of a direct dynamical model M , which describes the time evolution of the physical phenomenon. M allows connecting the geophysical variables studied with observations. By varying some geophysical variables (control variables); assimilation seeks to infer the physical variables that led to the observation values. These physical variables can be, for example, initial conditions or parameters of M .

The basic idea is to determine the minimum of a cost function J that measures the misfits between the observations and the model estimations. Due to the complexity of this function, the desired minimum is classically obtained by using gradient methods, which implies the use of the adjoint model of M . This model is derived from the equations of the direct model M . The adjoint model estimates changes in the control variables in response to a disturbance of the output values calculated by M . It is therefore necessary to proceed in the backward direction to the direct model calculations, which means to use the transpose of the Jacobean matrix. When observations are available, the adjoint allows minimizing the cost function J .

Formalism and notations for variational data assimilation are taken from Ide et al., (1997). M represents the direct model, $\mathbf{x}(t_0)$ is the initial state of the model, so $\mathbf{x}(t_i) = M_i(\mathbf{x}(t_0))$, where $M_i(\mathbf{x}(t_0))$ is represented by $M_1 M_2 M_3 \dots M_i(\mathbf{x}(t_0))$. The tangent linear model is noted as $\mathbf{M}(t_i, t_{i+1})$, which is the Jacobean matrix of M , in $\mathbf{x}(t_i)$. The background vector is defined as \mathbf{x}^b , which is an *a priori* state vector. \mathbf{y}_o^i is an observation at time step t_i . \mathbf{k} is a vector containing the control parameters of the model, so the direct model M can be represented also as $M_i(\mathbf{x}_i, \mathbf{k})$. The adjoint model \mathbf{M}_i^T is the linear tangent transpose, defined as:

$$\mathbf{M}_i^T(\mathbf{x}_i, \mathbf{k}) = \prod_{j=0}^{i-1} \mathbf{M}_i(t_j, t_{j+1})^T \quad (1)$$

M is used to estimate variables, which are most often observed from an observation operator H , allowing comparing the observed values and those calculated by the composition $H \circ M$, when they are available. The cost function J will be defined in terms of observations, so H_i allows us to compute observation variables \mathbf{y}_i , from state vector $\mathbf{x}(t_i)$. We suppose that $\mathbf{y}_i = H_i(M(\mathbf{x}_i, \mathbf{k})) + \varepsilon_i$, where ε_i is a random variable with zero mean. This term represents the sum of the model, observation and scaling error. Finally, the most general form of the cost function is defined as follows:

$$J(\mathbf{x}_0) = \frac{1}{2}(\mathbf{k}_0 - \mathbf{k}^b)B_0^{-1}(\mathbf{k}_0 - \mathbf{k}^b) + \frac{1}{2} \sum_{i=0}^n (\mathbf{y}_i - \mathbf{y}_0^i)R_0^{-1}(\mathbf{y}_i - \mathbf{y}_0^i) \quad (2)$$

The first part of the cost function represents the discrepancy to \mathbf{k}^b . The second part represents the distance between the observations and the model estimates. B_0 is the covariance error matrix of \mathbf{k}^b and R_i is the covariance error matrix of \mathbf{y}_o , at time t_i . The objective of this work is to show the capacity of 4DVAR to help determining the value of the principal parameters \mathbf{k} and initial conditions and also to investigate the impact of the prior choices. In our experiments, since the observations are synthetic (produced by the model itself) no transfer function from the estimation to the observation are needed, reason why we take H as the identity matrix. In addition since no *a priori* from the control parameters is known, there are no background, thus $B_0 = 0$;

If we note by $\nabla_{\mathbf{k}} J$ and $\nabla_{\mathbf{y}} J$ the gradients of the cost function J , with respect to \mathbf{k} and \mathbf{y}_i , we have:

$$\nabla_{\mathbf{k}} J = \mathbf{M}_i^T(\mathbf{x}_i, \mathbf{k}) \nabla_{\mathbf{y}} J \quad (3)$$

The expression above allows us to compute $\nabla_{\mathbf{k}} J$, by knowing $\nabla_{\mathbf{y}} J$, in the form of a matrix product of this term by the matrix $\mathbf{M}_i^T(\mathbf{x}_i, \mathbf{k})$, corresponding to the transpose of the Jacobian Matrix. The development of

calculation gives the expression of the gradient of y :

$$\nabla_k J = \mathbf{M}_i^T(x_i, k) \nabla_y J = \sum_{i=0}^t \mathbf{M}_i^T(x_i, k) H^T [\mathbf{R}_i^{-1} (y_i - y_i^o)] \quad (4)$$

Control parameters are adjusted several times until a stopping criterion is reached. The iterations of the gradient method allow us to approach the desired solution, in order to satisfy a stopping criterion that could be, for example, a certain threshold on the norm of the cost function gradient.

3.2 YAO

In order to implement the variational data assimilation, the gradient of the cost function need to be calculated before the minimization operation. YAO provides a framework helping the implementation of the adjoint model using a programming based on a general formalism decomposition of complex systems into modular graph (Nardi et al., 2009, <http://www.locean-ipsl.upmc.fr/~yao/>). The graph is composed of modules connected together by nodes and representing the numerical model. The module inputs are the outputs of precedent modules. Each module is composed of an elementary function specific to the dynamic model, which is differentiable. With YAO, the user specifies the type of discretization through a meta-language as well as the specifications of the direct model. YAO compiles and generates an executable that can compute the direct model M , the tangent linear model and the adjoint model. It also allows us to choose an implementation for the minimization of the cost function J , according to the specific scenario.

An interface with a quasi Newton optimizer called M1QN3 (Gilbert and Le Marechal 1989) is used to minimize the cost function, taking as parameters the cost function and its partial derivatives with respect to the control parameters. YAO, using forward and backward integrations of M and its adjoint, yields the derivatives. The stopping criterion for M1QN3 is defined by an epsilon, which is based on the norm of the current gradient g_k with respect to the initial gradient value g_1 . If $\|g_k / g_1\| < \epsilon$, the algorithm stops the minimization.

3.3 Experiment implementation

The implementation of SECHIBA in YAO starts with the definition of the modular graph describing the dynamics of the model M . Elementary processes and interconnections between modules are defined in order to grasp the essence of the model. This is a crucial step since the modular graph is the base for all the integration processes made by YAO. Direct and adjoint models are computed following the modular graph structure. The second step in the implementation outline is the coding of the direct and the derivatives of the modules. The process associated with each module was coded line-by-line, translating the original Fortran code in YAO meta-language, which is based on C++. At the end of the YAO implementation process, a validation of the direct model with respect to the original model was made. Once this validation was achieved, the next step of the implementation process is the derivative calculation of each module. This derivative process is based on the forward computing, in order to obtain the Jacobean matrix of each module. YAO generates automatically the tangent linear and the adjoint of the model.

Prior to the assimilation process, different scenarios are defined. A scenario makes reference to the experimental conditions. It includes the definition of: vegetation fraction (PFT), type of observation to be assimilated, observation sampling, time sampling, and atmospheric forcing file, subset of control parameters,

assimilation window size and wished time of the year to start the assimilation. The different scenarios were calculated using the adjoint model for several typical summer conditions of the 2 Fluxnet sites selected. The dates presented in this paper (2 February 2003 for Kruger Park and 26 August 1996 for Harvard Forest site) are representatives of sunny days in summer, with no perturbation coming from clouds and without rainfall events. All scenarios have normalized control parameters equal to 1.

The next section explains the scenarios for the different experiments performed in this work: a local sensitivity analysis using the adjoint model, a sensitivity cross-correlation between the gradients of temperature with respect to a set of control parameters and several twin experiments, based on the assimilation of simulated land surface temperature.

4. Data assimilation experiments

In order to show the benefit of data assimilation in SECHIBA, we conducted several experiments using the SECHIBA model. These experiments show the benefit in using data assimilation for estimating the internal parameters of the SECHIBA model. We first did a sensitivity analysis of SECHIBA to detect the most sensitive parameters (section 4.1 and 4.2); then, we conducted several data assimilation experiments using a twin experiment methodology to show the feasibility of estimating the internal parameters of the SECHIBA model (section 4.3).

4.1 Variational sensitivity analysis

In order to identify the most sensitive parameters to the estimation of land surface temperature, a sensitivity analysis was performed, based on the adjoint model of SECHIBA. The gradients obtained with the adjoint model correspond to first-order sensitivity, giving insights of the influence that the control parameters has on the land surface temperature. In order to do so, local sensitivities was applied, providing the slope of the calculated model output in the parameter space at a given set of values (Saltelli et al, 2008). This method is really local and the information provided is related to a single point in the parameter space. The point investigated is usually near the prior value of the parameters. Small variations in parameter values usually do not change the local sensitivity dramatically, but a significantly different parameter set may result in completely different sensitivity patterns (Saltelli et al., 2008). Local analysis has the advantage of facilitating the mechanistic interpretation of sensitivities (in sign and magnitude).

The accuracy of the adjoint of the model was tested as well; by comparing the gradients computed with the adjoint model to a gradients approximation, calculated using the direct model (finite differences). This comparison was made for every parameter considered in this study, allowing us to corroborate the precision of the adjoint model, verifying the exactitude of the approximated vs. the exact gradient values.

Parameters prior values used for the sensitivity analysis are the same as the original model ORCHIDEE. The sensitivity analysis is performed for a subset of internal parameters related to the energy and water physical processes, complementary to the work made by Kuppel et al. 2012, who considered only parameters related to the carbon budget.

The 11 parameters concerned in the analysis are presented in Table 1. There are two groups of parameters: internal parameters and multiplying factors. The first group corresponds to physical parameters. The second

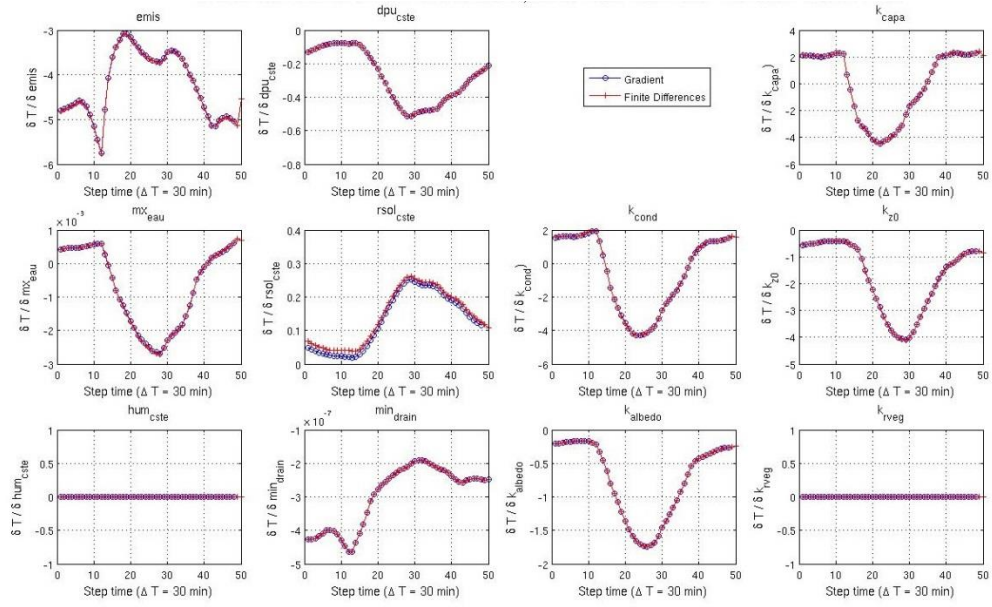
group resembles parameters weighting some physical processes in our model. All multiplying factor are equal to 1 so they do not influence the weighted equation in their prior values.

Model internal parameters are the following: $rsol_{cste}$ is a numerical constant involved in the soil resistance to evaporation. This parameter limits the soil evaporation, so the greater its value the lower the evaporation; hum_{cste} , mx_{eau} and min_{drain} are related to soil water processes, the higher their values, the more water will be available in the model reservoir, affecting water transfers and especially evapotranspiration; dpu_{cste} represents the soil depth in meters. The other parameters are multiplicative factors. We have k_{rveg} which is use in the calculation of the stomata resistance, this variable limits the transpiration capacity of leaves, the greater its value, the lower the transpiration; k_{emis} is the soil emissivity used to compute land surface temperature. This parameter takes part in the net radiation calculation which determines the balance between incoming and outgoing energy at the surface; k_{albedo} weights the surface albedo, which is defined as the reflection coefficient for short wave radiation; k_{cond} and k_{capa} take part in the thermal soil capacity and conductivity, both involved in the computation of the soil thermodynamics and k_{z0} weights the roughness height, which determines the surface turbulent fluxes. Although hum_{cste} is related to vegetation type, in this work only value for PFT 0 ($5 m^{-1}$) and PFT 11 ($2 m^{-1}$) are considered.

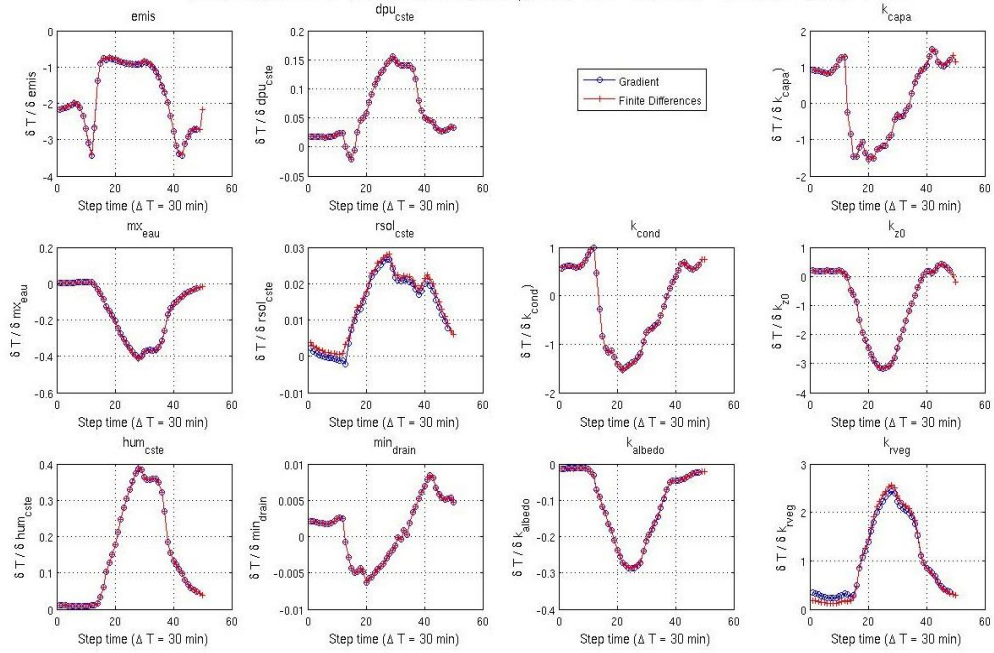
Parameter	Description	Prior Value	Unit
Inner Parameters			
hum_{cste}	Water stress	{5, 0.8, 0.8, 1, 0.8, 0.8, 1, 1, 0.8, 4, 4, 2, 4}	m^{-1}
$rsol_{cste}$	Evaporation resistance	33000	S/m^2
min_{drain}	Diffusion between reservoirs	0,001	-
dpu_{cste}	Total depth of soil water pool	2	m
mx_{eau}	Maximum water content	150	Kg/m^3
Multiplying Factors			
k_{emis}	Surface Emmisivity	1	-
k_{capa}	Soil Capacity	1	-
k_{cond}	Soil Conductivity	1	-
k_{rveg}	Vegetation Resistant	1	-
k_{z0}	Roughness height	1	-
k_{albedo}	Surface albedo	1	-

Table 1. SECHIBA Parameters studied in this work. There are 6 internal parameters, involved in the models' estimations and 5 multiplying factors that are imposed to specific fluxes

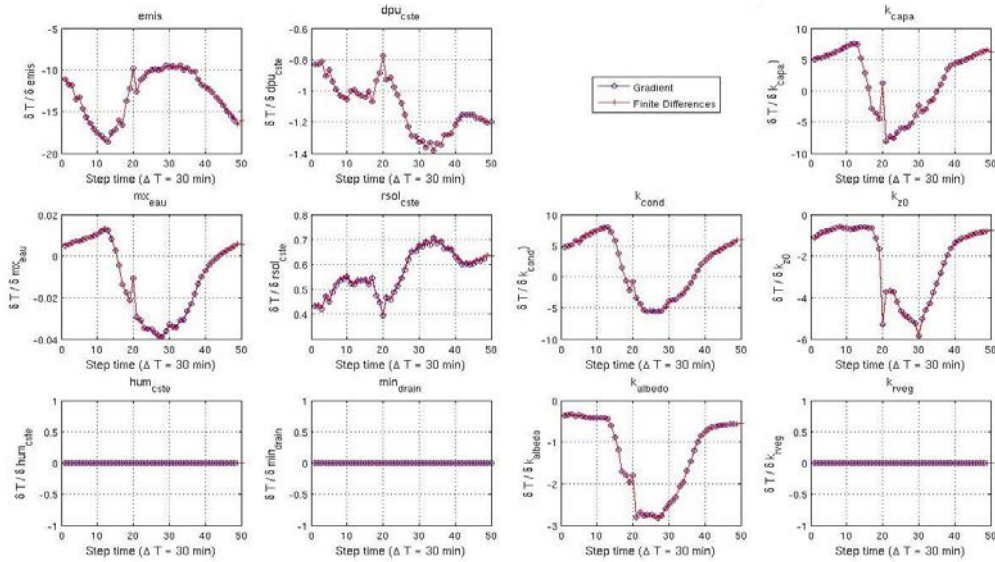
The sensitivity analysis was performed on bare soil (PFT 0) and grassland (PFT 11), in order to quantify the role of the vegetation on the land surface temperature parameters' sensitivity. The work is made on a daily basis, in order to observe the diurnal variations of sensitivities. At each half-hour time step, the model is restarted every time a gradient is computed, in order to have the updated gradient value.



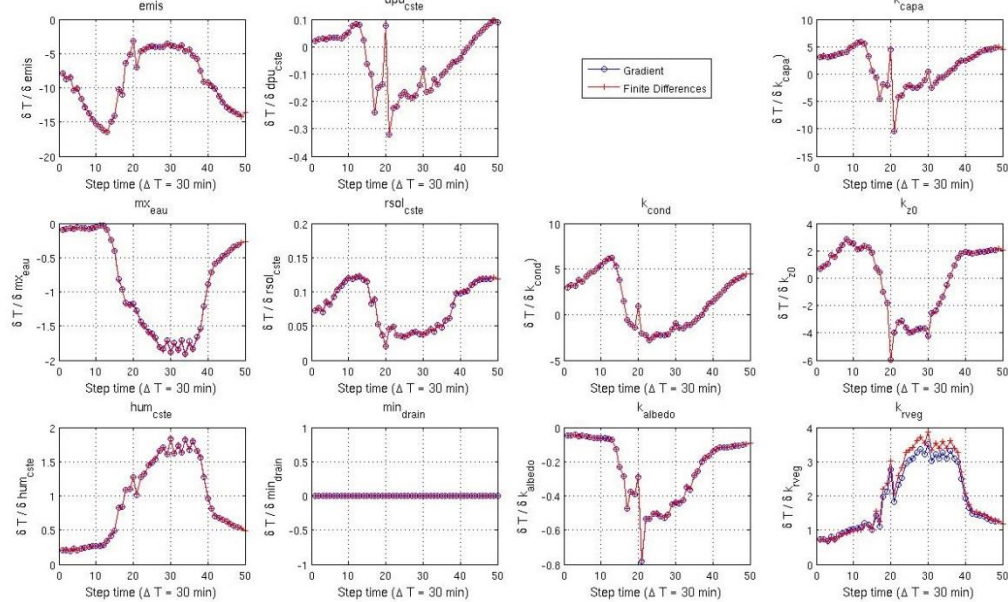
(a)



(b)



(c)



(d)

Figure 1. Sensitivity analysis results for PFT 0 and 11 using gradients obtained both with the adjoint model and finite differences. The sensitivities were computed on the surface temperature for Harvard Forest (Fig.1 (a) and (b)) and Kruger Park (Fig.1 (c) and (d)) sites. Blue curves represent the land surface temperature derivative with respect to each parameter given by the adjoint each half hour over a day. Red curves represent the temperature derivative computed with a finite difference discretization of the model. The first two columns correspond to model internal parameters. Third and fourth columns correspond to the multiplying factors added to the model.

Figure 1 compares the sensitivities obtained for each control parameter with both the finite differences and the model gradients. Bare soil results are presented in Fig.1a, (Harvard Forest) and Fig.1c (Kruger Park). The grassland scenario is illustrated in Fig.1b (Harvard Forest) and Fig.1d (Kruger Park). The efficiency of the

adjoint calculation is first demonstrated in these plots, because the 11 desired parameters sensitivities are obtained in a single integration. The results show little discrepancies in some gradients that should not affect the conclusion of our study.

Table 2 presents the 11 parameters ranked in order of their influence, according to the four scenarios defined. Parameter hierarchy revealed that the highest gradient values correspond to the ones that have the largest influence on the land surface temperature estimate. Clearly k_{emis} is the most influential parameter in the calculation of land surface temperature, regardless of the climatology used and vegetation fraction. In addition, min_{drain} is the least influential parameter for all defined scenarios.

Site	Bare Soil (PFT 0)	Grassland (PFT 11)
Harvard Forest	k_{emis} , k_{cond} , k_{capa} , k_{z0} , k_{albedo} , dpu_{cste} , $rsol_{cste}$, mx_{eau} , min_{drain} , k_{rveg} , hum_{cste}	k_{emis} , k_{rveg} , k_{cond} , k_{capa} , k_{z0} , mx_{eau} , hum_{cste} , k_{albedo} , dpu_{cste} , $rsol_{cste}$, min_{drain}
Kruger Park	k_{emis} , k_{cond} , k_{capa} , k_{z0} , k_{albedo} , dpu_{cste} , $rsol_{cste}$, mx_{eau} , min_{drain} , k_{rveg} , hum_{cste}	k_{emis} , k_{rveg} , k_{cond} , k_{capa} , k_{z0} , mx_{eau} , hum_{cste} , k_{albedo} , dpu_{cste} , $rsol_{cste}$, min_{drain}

Table 2. Parameter hierarchy according to each site and vegetation fraction.

The parameters k_{capa} , k_{cond} , k_{z0} and k_{albedo} are the most influential in bare soil conditions, after k_{emis} . In the presence of vegetation, several sensitivities change radically: k_{rveg} becomes the most important multiplicative factor after k_{emis} , k_{albedo} is less sensitive compared to its influence in the bare soil case and mx_{eau} is more sensitive, given that less water is available when a fraction of vegetation is present. The rest of the parameters show equivalent sensitivity values regardless the scenario. For hum_{cste} and k_{rveg} , sensitivities are equal to zero for bare soil, because these parameters affect surface temperature only in presence of vegetation.

Figure 1 also highlights the diurnal characteristics of the parameter sensitivities with a maximum around noon in line with the diurnal variation of solar radiation. Parameters with persistent positive sensitivity are: $rsol_{cste}$, k_{rveg} and hum_{cste} . Parameters with persistent negative sensitivity are: k_{z0} , k_{albedo} and $emis$. The sign of the gradients reflects the positive or negative feedback on the surface temperature of the processes involved. For example, the parameters involved in the evapotranspiration processes present negative sensitivities because a reduction (respectively increase) of the evapotranspiration will lead to an increase (respectively decrease) of the land surface temperature, when soil water content is sufficient. During the night, parameters sensitivities are equivalent for the different sites. In the case of bare soil, the model parameters present about the same sensitivities whatever the atmospheric forcing, but with a greater magnitude in Kruger Park, because the incident radiation is larger at this site.

Transpiration processes influence directly the land surface temperature in the presence of vegetation and is the dominant process in the studied sites. Therefore k_{rveg} has higher sensitivity than k_{cond} , k_{capa} and k_{albedo} . For bare soil, on the contrary, the dominant processes are those related to the soil thermodynamics, explaining why k_{capa} , k_{cond} and k_{emis} are the most sensitive parameters. In general, sensitivities are higher in bare soil conditions for the control parameters, except for min_{drain} and mx_{eau} .

In Kruger Park atmospheric forcing, the air temperature, air humidity and long wave radiation present an abrupt change at 10:00 AM (time step 20), due to a cold air stream event. This is reflected in the fluxes estimation and parameter sensitivities. In addition, when the difference between air temperature and land surface

temperature is maximum, the sensitivity is also maximum for k_{cond} and k_{capa} , and when both temperatures are equal; these two parameters have sensitivity equal to zero.

Furthermore, a sensitivity study was performed for different days in winter for both sites, in order to compare to the results obtained in summer. The same conditions were considered: two types of vegetation and clear sky days. The magnitude of the sensitivities in winter appeared lower than in summer, because of the weaker values of net radiation and surface fluxes but the same conclusions were obtained: same sign and diurnal variations of the sensitivities and same hierarchy of parameters.

We note that the derivatives provided by the adjoint and those given a finite difference approach of M are quasi-identical showing the ability of the adjoint to provide accurate estimates of the parameter derivatives.

4.2 Sensitivity cross-correlation

Pearson Correlation coefficients (r) were computed between land surface temperature gradients with respect to control parameters, during 4 days, from 2 February 2003, in Kruger Park with grassland, with the aim of identifying a subset of independent parameters and to document possible redundancies. Only 10 of the 11 parameters mentioned in Table 1 are taken into account since land surface temperature has a small sensitivity with respect to min_{drain} and thus this parameter is discarded.

Gradients were computed every 30 minutes, so we have a matrix of 192 gradient values for each land surface temperature gradients with respect to control parameters. The correlation coefficient r provides information on the degree of linear dependency between two gradients. A non-redundant parameter gradient has low correlation with the other gradients. It is determinable if land surface temperature is sensitive to it (high absolute normalized gradient value). The gradients were divided into two sets: those corresponding to daylight (between 7:00 and 20:00) where solar radiation is available and night (between 20:30 and 6:30) without solar radiation. The interest of splitting the correlation analysis into day and night comes from the need to study model dynamics. The processes implemented into the model vary widely during the day. By computing correlations, parameter dynamics arises, in line with the different processes that intervene during day and night.

Figure 2 presents the correlation matrix of the gradients for both time sub-periods. Upper triangular matrix corresponds to coefficients r computed with gradients during day time. Lower triangular matrix relates to coefficients computed with gradients during the night. Parameters were sorted according to the hierarchy obtained in Table 2. Strong correlations occur between several couples of gradients. These correlations are kept only if the gradient is greater than 1 Kelvin in that period. Cells with no color are correlations with p-value < 0.05 but with gradients under unity, thus they are excluded from the analysis. In general, if the gradients are small (cells with no-bold border) and the processes do not intervene in the period of study; the correlation is not interpretable in terms of physical processes, given the marginal influence these parameters have on land surface temperature variations. Some of these correlations are maintained during day and night (green background), meaning that the processes related to these parameters have the same influence in estimating of land surface temperature, independently of the hour of the day. On the other hand, blue cells present significant correlations at night (or day) only. The gray background cells correspond to correlations with p-values greater than to 0.05, indicating statically insignificant correlations.



redundant parameters, in the calculation of land surface temperature.

4.3 Twin experiments

Twin experiments are synthetic tests checking the robustness of the variational assimilation method. The model is run with a set of parameters or initial conditions ***Ptrue*** in order to produce pseudo observations of land surface temperature. Then ***Ptrue*** is randomly noised to obtain ***Pnoise***. Assimilations of land surface temperature were then performed during several days (most of the time, one week), leading to a new set of optimized parameters denoted ***Passim***. Four different assimilation experiments were performed.

Experiment Definition

The parameters considered in the twin experiments are taken from the sensitivity analysis results. Each experiment was perturbed with a uniform distribution random noise reaching 50% of the parameter prior value. Control parameters are normalized by their reference value so that relative perturbations are considered. If the control parameter values after the assimilation process are close to 1, it means that the parameter prior values were retrieved successfully. On the contrary, differences between the values retrieved and the prior value represent relative errors on the parameter estimation, after assimilation.

Conditions	Experiment 1	Experiment 2	Experiment 3	Experiment 4
Assimilation period	2 February 2003, 1 week (Kruger Park) 8 August 1996, 1 week (Harvard Forest)	2 February 2003, 1 week (Kruger Park) 8 August 1996, 1 week (Harvard Forest)	2 February 2003, 1 week (Kruger Park) 8 August 1996, 1 week (Harvard Forest)	2 February 2003, 4 days
Number of assimilations	500	500	500	500
Control Parameters	$k_{emis}, k_{cond}, k_{capa}, k_{z0}, k_{albedo}$	$k_{emis}, k_{rveg}, k_{cond}, k_{capa}, k_{z0}, k_{albedo}$	All parameters, except min_{drain}	Initial Surface Water Content
Observations	Land surface temperature	Land surface temperature	Land surface temperature	Land surface temperature
Observation sampling	30 minutes	30 minutes	30 minutes	30 minutes
Forcing	Kruger Park and Harvard Forest	Kruger Park and Harvard Forest	Kruger Park and Harvard Forest	Kruger Park
Vegetation type	PFT 0 (Bare Soil)	PFT 11 (Grassland)	PFT 11 (Grassland)	PFT 11 (Grassland)

Table 3. Scenarios for each of the 4 Experiments

Scenarios for all the assimilation experiments are presented in Table 3. All parameters are controlled at the same time. The duration of each assimilation experiment is one week and the time increment ΔT is 30 minutes. In Experiment 1 the five most sensitive parameters are controlled in bare soil conditions, during one week in Kruger Park and Harvard Forest sites. In Experiment 2 the sixth most sensitive parameters are controlled in conditions of grassland (PFT 11) in both FLUXNET sites during a week. With these two experiments, we are able to assess the effect of the vegetation fraction on the assimilation system. In addition, taking only the most sensitive parameters in the control set allow us to increase the assimilation performances, given that the more the observed variable is sensitive to a parameter, the easier the minimization process finds its optimal value, thus reducing the estimation error.

In Experiment 3, all parameters, except min_{drain} , are controlled (since min_{drain} has no impact in the land

surface temperature estimation), during a week, for both sites. Comparing Experiment 3 with Experiments 1 and 2 allows us to study the impact of taking a larger control parameter set in the assimilation process. In addition, we want to test if land surface temperature as observation, provide enough information to constrain the model parameters and if we can hope to improve all model state variables.

Experiment 4 controls the initial condition of the soil water content, during 4-days. In this experiment, no parameters were controlled. The idea is to find, after the optimization process the prior initial condition, which was perturbed prior to the assimilation.

We ran 500 assimilations in each experiment by randomly perturbing the initial conditions. This permitted us to obtain the control parameters relative errors and model fluxes root mean square error (RMSE) relative value, based on their value before and after the assimilation process. The fluxes considered are net radiation (R_n), soil heat flux (Q), sensible (H) and latent heat (LE) fluxes between the atmosphere and the biosphere, land surface temperature (LST) and the soil water reservoir contents (*Water stress*). The parameters considered depend on the experiment. The mean value of these statistics is presented in the different tables. With similar scenarios, Experiment 1 takes the 5 most sensitive parameters with bare soil, next the six most sensitive parameters are controlled with grassland for Experiment 2 and finally all the control parameter set is controlled for Experiment 3.

Harvard Forest		Experiment 1		Experiment 2		Experiment 3	
	Fluxes	Prior	Final	Prior	Final	Prior	Final
RMSE	Rn (W/m ²)	2.23.10 ⁻¹	1.27.10 ⁻¹⁹	2.31.10 ⁻¹	1.37.10 ⁻⁷	6.33.10 ⁻³	4.90.10 ⁻³
	Q(W/m ²)	1.32.10 ⁻¹	3.28.10 ⁻¹²	1.33.10 ⁻¹	2.86.10 ⁻⁶	2.31.10 ⁻¹	4.18.10 ⁻²
	H(W/m ²)	2.6.10 ⁻¹	8.47.10 ⁻¹²	2.95.10 ⁻¹	1.02.10 ⁻⁷	3.41.10 ⁻¹	4.99.10 ⁻²
	LE(W/m ²)	3.04.10 ⁻¹	1.84.10 ⁻¹⁷	3.81.10 ⁻¹	9.44.10 ⁻⁷	3.04.10 ⁻¹	1.84.10 ⁻²
	Water stress (%)	1.29.10 ⁻³	6.03.10 ⁻¹⁹	1.18.10 ⁻⁵	3.16.10 ⁻⁸	6.34.10 ⁻³	4.90.10 ⁻⁵
	Temperature (K)	1.67.10 ⁻²	5.51.10 ⁻¹⁵	1.35.10 ⁻²	7.66.10 ⁻⁶	5.12.10 ⁻²	1.01.10 ⁻⁴
		Prior	Final	Prior	Final	Prior	Final
Relative Error	Parameters						
	k _{emis}	2.55.10 ⁻¹	5.71.10 ⁻⁶	2.42.10 ⁻¹	5.96.10 ⁻⁷	2.63.10 ⁻¹	2.1.10 ⁻³
	k _{rveg}	-	-	2.21.10 ⁻¹	8.31.10 ⁻⁶	2.54.10 ⁻¹	1.79.10 ⁻²
	k _{cond}	2.41.10 ⁻¹	5.58.10 ⁻⁷	2.7.10 ⁻¹	5.96.10 ⁻⁶	2.51.10 ⁻¹	3.30.10 ⁻²
	k _{capa}	2.54.10 ⁻¹	5.57.10 ⁻⁸	2.69.10 ⁻¹	5.85.10 ⁻⁶	2.57.10 ⁻¹	2.61.10 ⁻²
	k _{z0}	2.44.10 ⁻¹	1.27.10 ⁻⁷	2.58.10 ⁻¹	7.84.10 ⁻⁷	2.57.10 ⁻¹	2.8.10 ⁻³
	k _{albedo}	2.44.10 ⁻¹	1.99.10 ⁻⁶	2.39.10 ⁻¹	2.08.10 ⁻⁶	2.47.10 ⁻¹	2.37.10 ⁻³
	m _x _{eau}	-	-	-	-	2.58.10 ⁻¹	7.34.10 ⁻²
	hum _{este}	-	-	-	-	2.52.10 ⁻¹	2.7.10 ⁻³
	dp _u _{este}	-	-	-	-	2.42.10 ⁻¹	2.2.10 ⁻³
	rsol _{este}	-	-	-	-	2.54.10 ⁻¹	2.36.10 ⁻³

Table 4. Model fluxes RMSE and Parameters Relative errors before and after the assimilation process for Experiment 3 to 5, on FLUXNET Harvard Forest

Results for Experiments 1, 2 and 3 are presented in Table 4, corresponding to Harvard Forest results, and Table 5, corresponding to Kruger Park site. In Experiment 1, the errors on the retrieved values for all the control parameters are of the order of 10⁻¹³. Regarding the land surface temperature, the mean RMSE ranges from 0.03, prior assimilation, to 3.10⁻¹⁴ after the assimilation process. Same behavior is observed for the different model fluxes. Experiment 2 yields similar results as in Experiment 1. The assimilation process allows reducing the parameter errors.

Kruger Park	Experiment 1	Experiment 2	Experiment 3
-------------	--------------	--------------	--------------

	Flux	Prior	Final	Prior	Final	Prior	Final
RMSE	Rn (W/m ²)	7.39.10 ⁻¹	2.6.10 ⁻²⁰	4.37.10 ⁻¹	2.28.10 ⁻⁸	1.13.10 ⁻¹	6.19.10 ⁻⁵
	Q(W/m ²)	1.60.10 ⁻¹	2.36.10 ⁻¹³	1.39.10 ⁻¹	3.21.10 ⁻⁷	1.88.10 ⁻¹	6.98.10 ⁻⁴
	H(W/m ²)	4.55.10 ⁻¹	8.47.10 ⁻¹³	4.73.10 ⁻¹	1.80.10 ⁻⁷	4.46.10 ⁻¹	1.46.10 ⁻³
	LE(W/m ²)	1.51.10 ⁻¹	8.84.10 ⁻¹⁷	1.24.10 ⁻¹	8.86.10 ⁻⁷	3.61.10 ⁻¹	1.01.10 ⁻²
	Water stress (%)	7.39.10 ⁻⁵	2.6.10 ⁻¹⁹	3.32.10 ⁻⁶	1.53.10 ⁻⁸	1.13.10 ⁻³	6.19.10 ⁻⁵
	Temperature (K)	3.41.10 ⁻²	1.62.10 ⁻¹⁵	2.75.10 ⁻²	1.51.10 ⁻⁷	1.54.10 ⁻²	3.4.10 ⁻⁶
	Parameters	Prior	Final	Prior	Final	Prior	Final
Relative Error	k _{emis}	2.58.10 ⁻¹	3.01.10 ⁻¹³	2.75.10 ⁻¹	6.08.10 ⁻⁷	2.41.10 ⁻¹	7.91.10 ⁻³
	k _{rveg}	-	-	2.81.10 ⁻¹	2.76.10 ⁻⁸	2.29.10 ⁻¹	4.91.10 ⁻³
	k _{cond}	2.54.10 ⁻¹	3.17.10 ⁻¹³	2.73.10 ⁻¹	6.37.10 ⁻⁸	2.38.10 ⁻¹	9.16.10 ⁻³
	k _{capa}	2.53.10 ⁻¹	3.1.10 ⁻¹³	2.73.10 ⁻¹	5.64.10 ⁻⁸	2.71.10 ⁻¹	7.86.10 ⁻³
	k _{z0}	2.51.10 ⁻¹	6.7.10 ⁻¹³	2.63.10 ⁻¹	7.97.10 ⁻⁷	2.43.10 ⁻¹	4.91.10 ⁻³
	k _{albedo}	2.59.10 ⁻¹	5.2.10 ⁻¹³	2.63.10 ⁻¹	2.31.10 ⁻⁶	2.53.10 ⁻¹	3.47.10 ⁻²
	mx _{eau}	-	-	-	-	2.46.10 ⁻¹	6.16.10 ⁻³
	hum _{este}	-	-	-	-	2.97.10 ⁻¹	3.7.10 ⁻²
	dpu _{este}	-	-	-	-	2.52.10 ⁻¹	2.6.10 ⁻²
	rsol _{este}	-	-	-	-	2.41.10 ⁻¹	1.26.10 ⁻²

Table 5. Model fluxes RMSE and Parameters Relative errors before and after the assimilation process for Experiment 3 to 5, on FLUXNET Kruger Park 02 February 2003

Several realizations of the prior parameters set in Experiments 1 and 2 (20% of the cases) did not converge at all. These results depend on the *a priori* parameter set value. They may indicate that a local minimum was reached in the minimization process of the cost function. In the case of non-convergence in Experiment 1, when the initial perturbations of k_{capa} and k_{cond} differ (in sign or magnitude), the assimilation process does not reduce the prior errors. As it was mentioned before, the correlation among k_{capa} and k_{cond} , affects the convergence in the parameter space. This highlights the difficulty to characterize these two parameters independently with only observations of land surface temperature.

Relative value of RMSE, with respect to synthetic measurements, for LE , H , Q and R_n in Experiment 3 prior to assimilation, are equal to 34%, 30% and 23% and 0.6%, respectively. After assimilation, RMSE is significantly reduced for both sites. The same holds for the mean relative error of the control parameters.

Comparing the results from Experiments 1 and 2 to Experiment 3, degradation in fluxes and parameter restitution can be observed. Indeed, we find higher errors in the fluxes and the final control parameters when increasing the size of the control parameter set. Best performances in the parameters restitution are always for the control of 5 parameters, except at the Harvard Forest site, when the same performances were obtained with 5 or 6 control parameters. When controlling k_{rveg} plus 5 parameters in Kruger Park, a degradation is observed. This can be explained by the nature of this site: a higher shortwave incident radiation involves a smaller stomata resistance, and a larger transpiration, amplifying the weight of this flux in the temperature estimate. k_{rveg} sensitivity depends on the site and the vegetation fraction. This is important when comparing Tables 4 and 5 with Table 6.

In Table 6, results of controlling the initial state of surface soil water content are presented. These results show good performances of the LST assimilation. The initial RMSE for land surface temperature goes from 10^{-2} down to 10^{-16} after assimilation process. Similar results were obtained for the sensible and latent heat fluxes, soil heat flux and net radiation. However, the influence that the initial water content has on the land surface temperature is limited in the studied cases; when perturbing its initial value, the prior RMSE remains very small for LST. Nonetheless, the different independent assimilation tests give conclusive results about the potential of

controlling the initial surface water content.

	Fluxes	Prior	Final
RMSE	Rn (W/m ²)	3.34.10 ⁻¹	3.33.10 ⁻⁶
	Q(W/m ²)	6.42.10 ⁻²	7.96.10 ⁻¹⁵
	H(W/m ²)	1.98.10 ⁻³	3.23.10 ⁻¹⁵
	LE(W/m ²)	7.8.10 ⁻³	5.76.10 ⁻¹⁶
	Water stress (%)	3.34.10 ⁻²	3.32.10 ⁻¹⁰
	Temperature (K)	1.74.10 ⁻²	1.09.10 ⁻¹⁶
	Initial Condition	Prior	Final
Relative Error	Surface water Content	3.25.10 ⁻¹	2.99.10 ⁻¹⁴

Table 6. Results for Experiment 4. Fluxes RMSE and Initial conditions for Surface Water Content Relative Error before and after the assimilation process, for FLUXNET Kruger Park, 02 February 2003, with PFT 11

4. Conclusion

In this study the adjoint of SECHIBA was implemented, using an adjoint semi-generator software called YAO. With SECHIBA adjoint, land surface temperature gradients with respect to each control parameter were computed, with the aim of carrying out a sensitivity analysis of the parameter influence on LST estimation. Gradients were also used to account for parameter correlations. Once the parameter hierarchy was set, twin experiments were performed for different scenarios, for testing the robustness of the assimilation scheme.

Land surface temperature assimilation has the potential of improving the parameters of LSM, by adjusting properly the control parameters and initial conditions. In a forecasting approach, this can be valuable, given that simulation can be more reliable and fitted to the actual measurements. The improvement in the model fluxes after the assimilation of LST was demonstrated. Twin experiments show the power of variational data assimilation to improve model parameter estimation. For different scenarios and forcing sites, the different experiments were successfully accomplished, meaning that a reduction in the fluxes errors was obtained by introducing information given by the LST synthetic observations. In addition, the influence that the size of the control parameter set has in the assimilation performance was shown.

Adding extra parameters to the control set increases the parameter space and the difficulty to minimize the cost function. Taking into consideration the results of assimilation of land surface temperature when controlling all parameter set (Experiment 3), we can see that, after having made several assimilation runs, land surface temperature does not provide enough information to constrain all parameter set, in order to improve the estimation of state variables in SECHIBA. In the case of controlling all parameters we cannot hope improving all model state variables.

The first contribution of this work to SECHIBA model is that the potential of LST variational data assimilation to improve SECHIBA estimations was shown. Assimilation with the YAO approach allows the implementation of different assimilation scenarios in a very simple way, when performing the different twin experiments: control parameters and observed variables (once the adjoint code has been generated), assimilation windows, observation sampling, time sampling and other different features can be changed easily.

The second contribution to SECHIBA is the sensitivity analysis results. They show exactly which parameters are the most sensitive in the model and serves as a guide to know which parameters have to be controlled during the assimilation process. However, it is important to mention that sensitivity analysis depends on the region, forcing, PFT, day and night cycle, among other factors. The correlation study highlighted the

interactions between parameters, giving insights into the model conceptualization, putting in evidence part of the model dynamics.

Acknowledgements

This work used eddy covariance data acquired by the FLUXNET community and in particular by the following networks: AmeriFlux (U.S. Department of Energy, Biological and Environmental Research, Terrestrial Carbon Program and AfriFlux). Dr. P. Peylin, F. Chevalier and M. Crépon are acknowledged for fruitful discussions. We thank also F. Maignan for its continuous support in the use of ORCHIDEE model, and Dr. M. Berrondo, for the assistance in writing this article.

5. References

Aubinet, M., Vesala, T., Papale, D. Eddy. Covariance: A Practical Guide to Measurement and Data Analysis. Springer Atmospheric Sciences Editions, United States of America. 2012.

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. *Bull. Am. Meteorol. Soc.*, 82, 2415–2434. 2001

Barrett, D., Renzullo, L. On the Efficacy of Combining Thermal and Microwave Satellite Data as Observational Constraints for Root-Zone Soil Moisture Estimation. *CSIRO Land and Water*, 1109–1127, Canberra, Australia. 2009.

Bateni, S.M., Entekhabi, D., Jeng, D.S. Variational assimilation of land surface temperature and the estimation of surface energy balance components. *Journal of Hydrology*, 481, 143–156. 2013.

Castelli F., Entekhabi, D., Caporali, E. Estimation of surface heat flux and an index of soil moisture using adjoint-state surface energy balance. *Water Resources Research*, 35, 10, 3115–3125. 1999.

d’Orgeval, T., Polcher, J., and Li, L. Uncertainties in modelling future hydrological change over west africa. *Climate Dynamics*, 26, 93–108. 2006.

Ducoudré, N., Laval, K., and Perrier, A. SECHIBA, a new set of parametrizations of the hydrologic exchanges at the land/atmosphere interface within the LMD atmospheric general circulation model. *J. Climate*, 6, 248–273. 1993.

Evensen, G. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, 53, 343–367. 2003.

Friedlingstein P., Joel G., Field C. B., Fung I. Toward an allocation scheme for global terrestrial carbon models. *Global Change Biology*, 5, 755–770. 1999.

Ghent, D., Kaduk, J., Remedios, J. and Balzter, H. Data assimilation into land surface models: The implications for climate feedbacks. *International Journal of Remote Sensing*, 3, 617 — 632. 2011.

Giering, R., Kaminski, T. Recipes for Adjoint Code Construction. *ACM Transactions on Mathematical Software*, 24, 437–474. 1998.

Gilbert, J.C., LeMaréchal, C. Some numerical experiments with variable-storage quasi Newton algorithms, *Maths. Program*, 45, 407–435. 1989.

Harrison, D.E., Chiodi A.M, Vecchi, G.A. Effects of surface forcing on the seasonal cycle of the eastern equatorial Pacific. *J. Mar. Res.* 67, 701–729. 2009.

Huang, C., Li, X., Lu, L. Retrieving land surface temperature profile by assimilating MODIS LST products with ensemble Kalman filter. *Cold and Arid Regions Environmental and Engineering Research Institute, CAS, Lanzhou, China*. 2003.

Ide, K., Courtier, P., Ghil, M. et Lorenc, A. Unified Notation for Data Assimilation : Operational, Sequential and Variational. Special Issue *J. Meteorological Society Japan*, 75, 181–189. 1997.

Kuppel, S., Peylin, P., Chevallier, F., Bacour, C., Maignan, F. and Richardson, A. Constraining a global ecosystem model with multi-site eddy-covariance data. *Biogeosciences*, 9, 3757–3776. 2012.

Lahoz, W; Khattatov, B. Data Assimilation Making Sense of Observations. Springer Editions. 2010.

Le Dimet, F.-X., Talagrand, O. Variational Algorithms for Analysis and Assimilation of Meteorological Observations: Theoretical Aspects. Dynamic Meteorology and Oceanography 38. 1986.

Nardi, L., Sorrow, C., Badran, F., and Thiria, S. *YAO: A Software for Variational Data Assimilation Using Numerical Models*. Computational Science and its Applications - ICCSA 2009. International Conference, 5593, 2, 621-636. 2009.

Pipunic, R. C., Walker, J. P., and Western, A. Assimilation of remotely sensed data for improved latent and sensible heat flux prediction: A comparative synthetic study. 19th International Congress on Modelling and Simulation, Perth, Australia. 2008.

Reichle, R., Walker, J., Koster, R., Houser, P. Extended versus Ensemble Kalman Filtering for Land Data Assimilation. Journal of Hydrometeorology, 3, 728-740. 2001.

Reichle, R., Kumar, S., Mahanama, S., Koster, R. D., and Liu, Q. Assimilation of satellite-derived skin temperature observations into land surface models. Journal of Hydrometeorology, 11, 1103-1122. 2010.

Ridler, M., Sandholt, I., Butts, M., Lerer, S., Mougin, E., Timouk, F., Kergoat, L., Madsen, H. Calibrating a soil-vegetation-atmosphere transfer model with remote sensing estimates of surface temperature and soil surface moisture in a semi-arid environment. Journal of Hydrology 436-437, 1-12. 2012.

Robert, C, Blayo, E., Verron, J. Comparison of reduced-order, sequential and variational data assimilation methods in the tropical Pacific Ocean. Ocean Dynamics 56, 5-6 (2006) 624-633. 2007

Saltelli, A. Sensitivity Analysis. John Wiley & Sons Edition. United States of America. 2008

Sitch, S., Smith, B., Prentice, I.C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J.O., Levis, S., Lucht, W., Sykes, M.T., Thonicke, K., Venevsky, S. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Glob. Change Biol. 9, 161 -185. 2003.

Conclusion of the article

In this work, the adjoint of SECHIBA was implemented, using an adjoint semi-generator software called YAO. With the SECHIBA adjoint, a sensitivity analysis was performed with respect to synthetic measurements of land surface temperatures. To do so, gradients of temperatures were computed with respect to each control parameter considered. These gradients were obtained with the adjoint model of SECHIBA. Once the hierarchy of the parameter influence in the estimation of LST was obtained, temperature gradients were also used to account for parameter correlations. Finally, twin experiments were performed for different scenarios, for testing the robustness of the assimilation scheme.

Results obtained suggest that land surface temperature assimilation has the potential of improving the parameters of LSM, by adjusting properly the control parameters and initial conditions. However, since we are in a twin experiment scheme, where the variations of the estimations are accounted only with synthetic data, no other source of error is introduced.

The improvement in the model variables after the assimilation of LST was demonstrated since the initial parameter set values were retrieved in most of the experiments. Twin experiments show the influence that the size of the control parameter set has in the assimilation performance. Adding extra parameters to the control set increases the parameter space and the difficulty of the minimization of the cost function. The potential of LST variational data assimilation to improve SECHIBA estimations was shown. Assimilation with the YAO approach allows the implementation of different assimilation scenarios in a very simple way, when performing the different twin experiments: control parameters and observed variables, assimilation windows, observation sampling, time sampling and other different features can be changed easily.

By using the sensitivity analysis, we obtain which parameters are the most sensitive in the model and a guide to know which parameters have to be controlled during the assimilation process. In addition, the correlation gave us insights to model conceptualization, enhancing our knowledge in the model dynamics.

From the previous experiments shown in this chapter, it was noticed that changing the value of parameters in the model does make some differences in the model behavior, depending on the parameter. The important thing to keep in mind is that sensitivity tests indicate that some parameter changes result in “greater,” or more significant, changes than others. In addition, different control parameter set values can change the path computed in the

modular graph representing the system dynamics.

The results of the sensitivity analysis presented in this Chapter are used to perform the experiments presented in Chapter 6. Several assimilation experiments of land surface temperature are performed with the purpose of optimizing the most sensitive parameters, using 2 FLUXNET forcing database.

Chapter 6

Twin Experiments with SECHIBA-YAO 1D using FLUXNET Measurements

6.1 Introduction

In this Chapter, twin experiments are implemented for optimizing the most sensitive parameters to the estimation of land surface temperature. This Chapter is a complement of the article presented in Chapter 5.

Twin experiments consist in applying assimilation procedures with simulated data of properties usually analogous to those of real data (Malanotte and Young, 1992; Miller and Cornuelle, 1999). They consist of synthetic tests checking the robustness of variational data assimilation. The results serve as indicators to determine if the assimilation scheme developed could work for real data. In the next section, experiment definition and results are presented for five experiments.

6.2 Experiment Definition

The objective of the experiments presented in this section is to test the effect the diverse scenarios have on the assimilation performances. A scenario is defined by several properties, described in Table 6.1.

Properties	Description
Assimilation period	Time window of the assimilation period.
Number of assimilations	For each experiment, a number of assimilation is made with the same scenario but with different control parameter first guesses
Control Parameters	Parameters to be optimized in the assimilation procedure
Observations	Model variables considered as observations
Observation sampling	Frequency sampling of the observations
Forcing	Data forcing used to perform the assimilation
Vegetation type	Vegetation fraction considered in the experiment

Table 6. 1 Scenario properties

In each experiment, control parameters are normalized by their prior value, so that only relative perturbations are considered. If the control parameter values posterior to the assimilation process are close to 1, it means that the parameter prior values were retrieved successfully. Differences between the values retrieved and the prior values represent relative errors on the parameter estimation, posterior to assimilation.

Scenarios for all the assimilation experiments are presented in Table 6.2. All parameters listed in each experiment are controlled at the same time. ΔT for the different experiments is set at 30 minutes. From experiments 1 to 5 both FLUXNET sites were used. In experiments 1 and 2, the six most sensitive parameters are controlled, with grassland vegetation. In Experiment 1 several observation samplings are tested going from 30 minutes up to 24 hours. During one month, five independent assimilation tests were run for each observation sampling. In Experiment 2, a weighted random noise is introduced in the observations, going from 10% up to 50% of the true value of the observation. Both Experiments 1 and 2 use constant perturbations of the control parameters (-50% its prior value) in order to assess the impact of varying the observation sampling and the noise in the observations.

Experiments 3 to 5 aim to test the parameter restitution by varying the size of the control parameter set. In Experiment 3, the five most sensitive parameters are controlled with bare soil, during one week using Kruger Park and Harvard Forest forcing. In Experiment 4, the six most sensitive parameters are controlled with grassland during a week. With these two experiments, we are able to compare the effect of optimizing k_{veg} on the assimilation system. In Experiment 5, all parameters, except min_{drain} , are controlled (since min_{drain} has no impact on LST estimation). Comparing Experiment 5 with Experiments 3 and 4 allow us to study the impact of taking a larger control parameter set in the assimilation process.

Scenario	Parameters				
	Experiment 1	Experiment 2	Experiment 3	Experiment 4	Experiment 5
Assimilation period	11/02/2003, 1 month (Kruger Park) 28/08/1996, 1 month (Harvard Forest)	11/02/2003, 1 week (Kruger Park) 28/08/1996, 1 week (Harvard Forest)	11/02/2003, 1 week (Kruger Park) 28/08/1996, 1 week (Harvard Forest)	11/02/2003, 1 week (Kruger Park) 28/08/1996, 1 week (Harvard Forest)	11/02/2003, 1 week (Kruger Park) 28/08/1996, 1 week (Harvard Forest)
Number of assimilations	5 experiments 5 for each site	500	500	500	500
Control Parameters	$k_{rveg}, k_{capa}, k_{cond}, k_{zo}, k_{albedo}, k_{emis}$	$k_{rveg}, k_{capa}, k_{cond}, k_{zo}, k_{albedo}, k_{emis}$	$k_{capa}, k_{cond}, k_{zo}, k_{albedo}, k_{emis}$	$k_{rveg}, k_{capa}, k_{cond}, k_{zo}, k_{albedo}, k_{emis}$	All parameters, except min_{drain}
Observations	Soil Temperature	Soil Temperature with noise	Soil Temperature	Soil Temperature	Soil Temperature
Observation sampling	30 minutes, 2, 6, 12 and 24 hours	1 hour	30 minutes	30 minutes	30 minutes
Forcing	Kruger Park and Harvard Forest	Kruger Park and Harvard Forest	Kruger Park and Harvard Forest	Kruger Park and Harvard Forest	Kruger Park and Harvard Forest
Vegetation type	PFT 11 (Grassland)	PFT 11 (Grassland)	PFT 0 (Bare Soil)	PFT 11 (Grassland)	PFT 11 (Grassland)

Table 6. 2 Scenarios for each of the 6 Experiments.

6.3. Results

For each of the five hundred assimilations in Experiments 2 to 5, and the five assimilations in Experiment 1, we assimilate synthetic observations of land surface temperature. Performances are defined as the comparison between prior and final control parameter relative errors and model variables root mean square error (RMSE). These variables are computed in the same period as the synthetic observations. The variables considered are H , LE , Rn , Q and LST . The mean value of these statistics is presented in the different Tables.

6.3.1 Effect of the observation sampling

Experiment 1 investigates the impact of the observation sampling in the assimilation process, since varying the observation frequency lead to different amount of observations available for each assimilation experiment.

Each test in this experiment was labeled with a number. This number serves as reference to compare the different results. In Test 4, only two observations per day are taken at noon and at midnight. In Test 5, we have one observation per day, taken at noon.

Test Number	Sampling Frequencies	Observations per day	Observation per month
1	30 minutes	48	1440
2	2 hours	24	720
3	6 hours	4	120
4	12 hours	2	60
5	24 hours	1	30

Table 6. 3 Sampling frequencies for Experiment 1

Kruger Park			# Test (Posterior to assimilation)				
	Variables	Prior	1	2	3	4	5
RMSE	Rn (W/m^2)	18.4	0.07	0.407	0.504	5.99	10.3
	Q (W/m^2)	35	0.049	0.25	0.453	4.57	6.72
	H (W/m^2)	25	0.437	0.138	0.43	4.7	10.34
	LE (W/m^2)	15.7	0.0601	0.592	0.594	2.43	10.8
	$Water\ stress$ (m^3/m^2)	$2.5 \cdot 10^{-2}$	$2.57 \cdot 10^{-3}$	$3.95 \cdot 10^{-3}$	$1.57 \cdot 10^{-2}$	$1.8 \cdot 10^{-2}$	$2.7 \cdot 10^{-2}$
	LST (K)	7.98	0.0601	0.0243	0.592	0.594	1.9
	Parameters	Prior	1	2	3	4	5
Relative Error (%)	k_{cond}	50	0.0183	0.261	0.340	0.921	4.96
	k_{capa}	50	0.0427	0.172	0.4006	0.91	3.77
	k_{z0}	50	0.00103	0.0162	0.147	0.24	1.34
	k_{rveg}	50	0.418	0.909	3.845	4.01	14.97
	k_{emis}	50	0.1704	0.2733	0.77	1.27	4.4
	k_{albedo}	50	0.128	1.384	3.214	4.15	25.01

Table 6. 4 Variable RMSE and control parameter relative errors prior and posterior to the assimilation, for Experiment 1 (different frequencies sampling in the observations), using Kruger Park forcing, PFT 11, from 11/02/2003 to 11/03/2003.

Prior and final errors before and posterior to the assimilation process are presented in Table 6.4 and 6.5 for Kruger Park and Harvard Forest sites, respectively. The columns represent the different assimilations performed with different frequency sampling in the observations. Five independent assimilations were made. The Tables report the mean value of the performances of the assimilation system. Even though small errors were found for the different tests, we do notice that the assimilation system is sensitive to the observation sampling.

The contribution from the data is demonstrated by an improvement in the optimization when increasing the frequency of observation. The final error values in the different tests increase by a factor of 10 when reducing the sampling frequency. By decreasing the number of observations, the control parameters adjustment is less accurate, and the model estimate variables with a larger error. Thus it can be verified that if we have more observations of LST, the assimilation system will fit the parameters better, so improved estimations are obtained.

Harvard Forest			# Test (Posterior to assimilation)				
	Variables	Prior	1	2	3	4	5
RMSE	$Rn (W/m^2)$	71.59	1.75	3.16	22.32	48.82	49.05
	$Q (W/m^2)$	21.79	0.19	0.91	3.13	5.63	5.65
	$H (W/m^2)$	13.42	0.15	0.98	1.84	3.98	4.08
	$LE (W/m^2)$	86.23	0.22	0.35	3.81	5.17	11.95
	$Water\ stress$ (m^3/m^3)	$1.7 \cdot 10^{-2}$	$6.7 \cdot 10^{-3}$	$1.5 \cdot 10^{-3}$	$1.65 \cdot 10^{-2}$	$1.69 \cdot 10^{-2}$	$2.1 \cdot 10^{-2}$
	$LST (K)$	5.98	0.08	0.65	0.86	1.27	1.61
			1	2	3	4	5
Relative Error (%)	Parameters	Prior					
	k_{cond}	50	0.37	0.54	3.7	5.7	10.14
	k_{capa}	50	0.36	2.86	4.16	10.55	20.74
	k_{z0}	50	0.0592	0.15	7.61	13.74	16.73
	k_{rveg}	50	0.31	0.75	5.25	7.24	17.8
	k_{emis}	50	0.11	0.17	5.82	10.86	13.74
	k_{albedo}	50	1.54	4.81	12.69	34.11	37.8

Table 6. 5 Variable RMSE and control parameter relative errors prior and posterior to assimilation, for Experiment 1 (different frequencies sampling in the observations), using Harvard Forest forcing, PFT 11, from 28/08/1996 to 28/09/1996.

6.3.2 Effect of random noise in the observation

Experiment 2 aims to study the impact of introducing a random noise in the synthetic observations. The random noise follows a normal distribution with zero mean and variance 1. The perturbed observations are computed using the following equation

$$LST^* = LST + amp \cdot \phi \quad (6.1)$$

with LST^* the perturbed observation, LST the original land surface temperature, amp a factor weighting the random noise going from 10% to 50%, and ϕ is the normal distribution random noise. The control parameter set is composed of the six most influential parameters in the computation of LST . The first guess is obtained by perturbing 10% uniformly the control parameter set from its prior values. Three experiments are performed, aiming to test the impact of introducing different magnitudes of errors prior the assimilation process. The mean value of the five hundred independent assimilations is presented. Posterior to each experiment, the parameter relative error and the model flux RMSE are computed to quantify the quality of the results.

We note in Tables 6.6 and 6.7 that the parameter restitution is degraded when adding random noise to the observations. This shows the sensitivity of the assimilation system has to noise affecting the observations of LST .

		Kruger Park		Amplitude		
		Variables	Prior	10%	30%	50%
RMSE	Rn (W/m^2)	20.67	9.9	9.96	10.09	
	Q (W/m^2)	4.24	1.88	2.07	3.74	
	H (W/m^2)	24.7	7.26	7.81	8.32	
	LE (W/m^2)	4.06	3.78	3.9	6.22	
	$Water\ stress$ (m^3/m^3)	$2.1 \cdot 10^{-2}$	$1.5 \cdot 10^{-2}$	$1.7 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$	
	LST (K)	7.12	0.019	4.48	6.23	
Relative Error (%)	Parameters	Prior	10%	30%	50%	
	k_{cond}	10	4.5	4.9	11.12	
	k_{capa}	10	1.51	3.35	14.9	
	k_{z0}	10	3.24	3.99	4.09	
	k_{rveg}	10	6.91	10.1	11.5	
	k_{emis}	10	2.79	3.14	4.08	
	k_{albedo}	10	1.12	2.01	3.02	

Table 6. 6 Variables RMSE and control parameter relative errors prior and posterior to assimilation, for Experiment 2 (different amplitude of random noise in the observation), using Kruger Park forcing, from 11/02/2003 to 18/02/2003.

		Harvard Forest		Amplitude		
		Variables	Prior	10%	30%	50%
RMSE	Rn (W/m^2)	18.4		5.80	5.95	6.01
	Q (W/m^2)	35		3.97	9.58	10.23
	H (W/m^2)	25		5.92	11.13	24.01
	LE (W/m^2)	15.7		4.77	14.04	15.05
	$Water\ stress$ (m^3/m^3)	$6.1 \cdot 10^{-2}$		$2.3 \cdot 10^{-2}$	$4.5 \cdot 10^{-2}$	$8.1 \cdot 10^{-2}$
	LST (K)	7.98		0.046	1.42	2.59
Relative Error (%)	Parameters	Prior	1	2	3	
	k_{cond}	10	0.83	4.32	7.6	
	k_{capa}	10	4.47	9.05	9.21	
	k_{z0}	10	3.85	4.5	7.3	
	k_{rveg}	10	1.36	7.01	8.04	
	k_{emis}	10	2.39	3.62	6.47	
	k_{albedo}	10	1.02	2.58	7.85	

Table 6. 7 Variables RMSE and control parameter relative errors prior and posterior to assimilation, for Experiment 2 (different amplitude of random noise in the observation), using Harvard Forest forcing, from 28/08/1996 to 04/09/1996.

When increasing the amplitude of the error, the various errors obtained for the three tests not only suggest the need to take into account the quality of the observations for the model but also the fact that the parameters are not affected in the same way by the data uncertainties. However, perturbations are still limited and a deeper exploration should be performed to assess the true impact in the assimilation performance of noisy observations.

6.3.3 Effect of the control parameter set size

Experiments 3 to 5 test the effect of changing the size of the control parameter set, in optimizing different parameters. In each experiment a random noise is added to the parameter

prior values. This random noise is computed independently for each parameter.

Comparison between each of these experiments evaluates the impact of varying the size of the control parameter set in the assimilation performance. With similar scenarios, Experiment 3 takes the five most sensitive parameters with bare soil, next the six most sensitive parameters are controlled with grassland for Experiment 4, and finally all the control parameter set is controlled for Experiment 5, with grassland.

Kruger Park		Experiment 3		Experiment 4		Experiment 5	
Variables		Prior	Final	Prior	Final	Prior	Final
RMSE	$Rn (W/m^2)$	$7.39 \cdot 10^{-1}$	$2.6 \cdot 10^{-20}$	$4.37 \cdot 10^{-1}$	$2.28 \cdot 10^{-8}$	$1.13 \cdot 10^{-1}$	$6.19 \cdot 10^{-5}$
	$Q (W/m^2)$	$1.60 \cdot 10^{-1}$	$2.36 \cdot 10^{-13}$	$1.39 \cdot 10^{-1}$	$3.21 \cdot 10^{-7}$	$1.88 \cdot 10^{-1}$	$6.98 \cdot 10^{-4}$
	$H (W/m^2)$	$4.55 \cdot 10^{-1}$	$8.47 \cdot 10^{-13}$	$4.73 \cdot 10^{-1}$	$1.80 \cdot 10^{-7}$	$4.46 \cdot 10^{-1}$	$1.46 \cdot 10^{-3}$
	$LE (W/m^2)$	$1.51 \cdot 10^{-1}$	$8.84 \cdot 10^{-17}$	$1.24 \cdot 10^{-1}$	$8.86 \cdot 10^{-7}$	$3.61 \cdot 10^{-1}$	$1.01 \cdot 10^{-2}$
	$Water\ stress$ (m^3/m^3)	$7.39 \cdot 10^{-2}$	$2.6 \cdot 10^{-19}$	$3.32 \cdot 10^{-2}$	$1.53 \cdot 10^{-8}$	$1.13 \cdot 10^{-2}$	$6.19 \cdot 10^{-5}$
	$LST (K)$	$3.41 \cdot 10^{-2}$	$1.62 \cdot 10^{-15}$	$2.75 \cdot 10^{-2}$	$1.51 \cdot 10^{-7}$	$1.54 \cdot 10^{-2}$	$3.4 \cdot 10^{-6}$
Relative Error	Parameters	Prior	Final	Prior	Final	Prior	Final
	k_{emis}	$2.58 \cdot 10^{-1}$	$3.01 \cdot 10^{-13}$	$2.75 \cdot 10^{-1}$	$6.08 \cdot 10^{-7}$	$2.41 \cdot 10^{-1}$	$7.91 \cdot 10^{-3}$
	k_{rveg}	-	-	$2.81 \cdot 10^{-1}$	$2.76 \cdot 10^{-8}$	$2.29 \cdot 10^{-1}$	$4.91 \cdot 10^{-3}$
	k_{cond}	$2.54 \cdot 10^{-1}$	$3.17 \cdot 10^{-13}$	$2.73 \cdot 10^{-1}$	$6.37 \cdot 10^{-8}$	$2.38 \cdot 10^{-1}$	$9.16 \cdot 10^{-3}$
	k_{capa}	$2.53 \cdot 10^{-1}$	$3.1 \cdot 10^{-13}$	$2.73 \cdot 10^{-1}$	$5.64 \cdot 10^{-8}$	$2.71 \cdot 10^{-1}$	$7.86 \cdot 10^{-3}$
	k_{z0}	$2.51 \cdot 10^{-1}$	$6.7 \cdot 10^{-13}$	$2.63 \cdot 10^{-1}$	$7.97 \cdot 10^{-7}$	$2.43 \cdot 10^{-1}$	$4.91 \cdot 10^{-3}$
	k_{albedo}	$2.59 \cdot 10^{-1}$	$5.2 \cdot 10^{-13}$	$2.63 \cdot 10^{-1}$	$2.31 \cdot 10^{-6}$	$2.53 \cdot 10^{-1}$	$3.47 \cdot 10^{-2}$
	mx_{eau}	-	-	-	-	$2.46 \cdot 10^{-1}$	$6.16 \cdot 10^{-3}$
	hum_{cste}	-	-	-	-	$2.97 \cdot 10^{-1}$	$3.7 \cdot 10^{-2}$
	dpu_{cste}	-	-	-	-	$2.52 \cdot 10^{-1}$	$2.6 \cdot 10^{-2}$
	$rsol_{cste}$	-	-	-	-	$2.41 \cdot 10^{-1}$	$1.26 \cdot 10^{-2}$

Table 6. 8 Surface variables RMSE and parameter relative errors before and posterior to the assimilation process for Experiment 3 to 5 (different control parameter set), using Kruger Park forcing, from 11/02/2003 to 18/02/2003.

Results for Experiments 3, 4 and 5 are presented in Tables 6.8 and 6.9. In Experiment 3, the errors on the retrieved values for the whole control parameter set are of the order of 10^{-13} . Regarding LST , the mean RMSE ranges from 0.03, prior assimilation, to $3 \cdot 10^{-14}$ posterior to the assimilation process. We observe the same behavior for the different model variables. Experiment 4 yields similar results as Experiment 3. The assimilation process allows reducing the parameter errors.

However, we should note that several estimations of the prior parameter set in Experiment 3 and 4 (20% of the cases) did not converge at all. These results depend on the prior parameter set values. The reason is probably that several independent assimilations should be conducted in order to avoid these types of results. They may indicate that a local minimum was reached in the minimization process of the cost function. In the case of non-convergence in Experiment 3, when the perturbations of k_{capa} and k_{cond} differ (in sign or

magnitude) the assimilation process does not reduce its initial error.

Harvard Forest		Experiment 3		Experiment 4		Experiment 5	
Variables		Prior	Final	Prior	Final	Prior	Final
RMSE	R_n (W/m^2)	$2.23 \cdot 10^{-1}$	$1.27 \cdot 10^{-19}$	$2.31 \cdot 10^{-1}$	$1.37 \cdot 10^{-7}$	$6.33 \cdot 10^{-3}$	$4.90 \cdot 10^{-5}$
	Q (W/m^2)	$1.32 \cdot 10^{-1}$	$3.28 \cdot 10^{-12}$	$1.33 \cdot 10^{-1}$	$2.86 \cdot 10^{-6}$	$2.31 \cdot 10^{-1}$	$4.18 \cdot 10^{-2}$
	H (W/m^2)	$2.6 \cdot 10^{-1}$	$8.47 \cdot 10^{-12}$	$2.95 \cdot 10^{-1}$	$1.02 \cdot 10^{-7}$	$3.41 \cdot 10^{-1}$	$4.99 \cdot 10^{-2}$
	LE (W/m^2)	$3.04 \cdot 10^{-1}$	$1.84 \cdot 10^{-17}$	$3.81 \cdot 10^{-1}$	$9.44 \cdot 10^{-7}$	$3.04 \cdot 10^{-1}$	$1.84 \cdot 10^{-2}$
	Water stress (m^3/m^3)	$1.29 \cdot 10^{-2}$	$6.03 \cdot 10^{-19}$	$1.18 \cdot 10^{-2}$	$3.16 \cdot 10^{-8}$	$6.34 \cdot 10^{-2}$	$4.90 \cdot 10^{-5}$
	LST (K)	$1.67 \cdot 10^{-2}$	$5.51 \cdot 10^{-15}$	$1.35 \cdot 10^{-2}$	$7.66 \cdot 10^{-6}$	$5.12 \cdot 10^{-2}$	$1.01 \cdot 10^{-4}$
Relative Error	Parameters	Prior	Final	Prior	Final	Prior	Final
	k_{emis}	$2.55 \cdot 10^{-1}$	$5.71 \cdot 10^{-6}$	$2.42 \cdot 10^{-1}$	$5.96 \cdot 10^{-7}$	$2.6 \cdot 10^{-1}$	$2.1 \cdot 10^{-3}$
	k_{rveg}	-	-	$2.21 \cdot 10^{-1}$	$8.31 \cdot 10^{-6}$	$2.5 \cdot 10^{-1}$	$1.79 \cdot 10^{-2}$
	k_{cond}	$2.41 \cdot 10^{-1}$	$5.58 \cdot 10^{-7}$	$2.7 \cdot 10^{-1}$	$5.96 \cdot 10^{-6}$	$2.1 \cdot 10^{-1}$	$3.30 \cdot 10^{-2}$
	k_{capa}	$2.54 \cdot 10^{-1}$	$5.57 \cdot 10^{-8}$	$2.69 \cdot 10^{-1}$	$5.85 \cdot 10^{-6}$	$2.7 \cdot 10^{-1}$	$2.61 \cdot 10^{-2}$
	k_{z0}	$2.44 \cdot 10^{-1}$	$1.27 \cdot 10^{-7}$	$2.58 \cdot 10^{-1}$	$7.84 \cdot 10^{-7}$	$2.5 \cdot 10^{-1}$	$2.8 \cdot 10^{-3}$
	k_{albedo}	$2.44 \cdot 10^{-1}$	$1.99 \cdot 10^{-6}$	$2.39 \cdot 10^{-1}$	$2.08 \cdot 10^{-6}$	$2.4 \cdot 10^{-1}$	$2.37 \cdot 10^{-3}$
	mx_{eau}	-	-	-	-	$2.8 \cdot 10^{-1}$	$7.34 \cdot 10^{-2}$
	hum_{cste}	-	-	-	-	$2.5 \cdot 10^{-1}$	$2.7 \cdot 10^{-3}$
	dpu_{cste}	-	-	-	-	$2.4 \cdot 10^{-1}$	$2.2 \cdot 10^{-3}$
	$rsol_{cste}$	-	-	-	-	$2.5 \cdot 10^{-1}$	$2.36 \cdot 10^{-3}$

Table 6. 9 Surface variable RMSE and parameter relative errors before and posterior to the assimilation process for Experiments 3 to 5 (different control parameter set), using Harvard Forest forcing, from 28/08/1996 to 28/08/1996.

RMSE values in Experiment 5 for LE , H , Q and R_n prior to assimilation are of the order of 34%, 30% and 23% and 0.6%, respectively. Posterior to assimilation, RMSE is reduced for both sites, as well as for the mean relative error of the control parameters.

Comparing the results from Experiments 4 and 5, degradation in variables and parameter restitution can be observed. Indeed, we find larger errors in the variable estimation when increasing the control parameter set from six to ten parameters. The best performance in the parameters restitution is always found when controlling six parameters.

In Experiment 4, some degradation is observed, when controlling k_{rveg} plus five parameters in Kruger Park. This can be explained by the nature of this site: a higher shortwave incident radiation involves a smaller stomatal resistance, increasing transpiration, amplifying the weight of this flux in the temperature estimation. We deduce that it is better to control only the common coefficients where performance is best. k_{rveg} sensitivity depends on the site and the vegetation fraction.

Controlling only the most sensitive parameters to LST increases our chances to find acceptable control parameter values after assimilation. Optimizing a larger control parameter set, as in Experiment 6, makes more difficult for the assimilation system to retrieve the prior value of the control parameters with higher accuracy.

6.4. Discussion

This chapter is intended as a complement to the results obtained in the previous Chapter 5, when the most sensitive parameters to LST were controlled in twin experiments. The results reaffirm the potential of 4DVAR in the optimization of SECHIBA parameters, in order to improve variables estimation.

By performing twin experiments, we can have an idea of the effect the different assimilation scenarios have on the parameter restitution performances. In addition, it allows us to study the impact the assimilation has in the improvement of model flux estimations. After presenting the different experiments, some aspects of data assimilation arise when analyzing the results. The first one concerns the presence of several local minima due to the non-linearity of the SECHIBA model. Second, we have also shown a significant improvement in the assimilation performances when the sampling frequency of observations is increased, as evaluated in Experiment 1. This suggests that the ability of the model to be constrained depends, among other things, on the observations frequency.

Finally, we observe a strong dependence between the quality of observations and the parameters restitution, as in Experiment 2. It seems crucial to take into account the uncertainty in the observations, because they do not affect the assimilation performance in the same way when estimating each parameter in the minimization process. The introduction of a regularization term on the parameters could be used to mitigate this problem. Constraining parameters and weighing observations according to their confidence in the minimization phase can be modeled through the introduction in the cost function of the variance-covariance errors matrices terms (background \mathbf{B} and observation \mathbf{R}). Of course, this is out of the scope of the twin experiments, but it is an important aspect to consider with real observations.

Once twin experiments using FLUXNET forcing were evaluated, the next step is to investigate the impact of VDA in the estimation of parameters with actual measurements, using several observed variables. By following the same methodology followed throughout Chapter 5 and 6, Chapter 7 presents the results of assimilating LST measurements in the framework of SMOSREX experimental database.

Chapter 7

Real measurements study using SMOSREX dataset

7.1 Introduction

In this Chapter, 4DVAR is applied using real measurements from SMOSREX site, as presented in Chapter 2. This dataset includes measurements of the estimated variables computed with SECHIBA-YAO 1D, so they can potentially be used as observations in the assimilation scheme, or they can be used to assess the improvements or degradation obtained after assimilating a specific measurement.

Previous works show the potential of constraining model parameter via the assimilation of brightness temperature measured in SMOSREX. In their work, Parrens et al., (2013) applied a simplified extended Kalman Filter (SEKF) in the surface model ISBA-DF, to assimilate microwave brightness temperature observations from the site SMOSREX, in order to better constrain the water content of the soil.

The guiding principle of this Chapter is to assess the impact of assimilating land surface temperature for constraining model parameters, based on assimilation performances. The assimilation methodology is based on the optimization of the most sensitive parameters to *LST*, as in Chapter 6. The experiments controlling the six and respectively ten most sensitive parameters are presented, in order to test the effect of changing the control parameter set. Assimilation performances are analyzed by taking a one-day and a seven-day assimilation time window. The daily-base assimilations are done during a month, with the idea of accounting for parameters variability.

For each assimilation experiment presented, variable root mean square errors (RMSE) are computed based on variable estimation with prior and posterior values of the control

parameter set. The final objective of the *LST* assimilation being the improvement of surface processes modeling, the variables considered to assess assimilation contribution are the sensible heat flux (H), latent heat flux (LE), net radiation (R_n), soil moisture (θ_s) and land surface temperature (LST). Since SMOSREX dataset includes measurements of these variables, the prior and posterior RMSE can be computed.

7.2 Key parameters to perform the optimization

The different experiments presented in this Chapter aims at optimizing the SECHIBA parameters listed in Table 7.1. The prior range of variation is prescribed, defining the interval in which each parameter is considered physically valid. The equations where each parameter occurs are presented in Chapter 2.

For the six multiplying factors, their prior value is 1, which means that they do not affect the corresponding model variables they are weighting. Their prior range is $\pm 50\%$, i.e. their prior value is between 0.5 and 1.5. In addition, the other five model internal parameters were normalized such as their prior value is 1. The internal parameters interval correspond to $\pm 50\%$ their true value for all the parameters, excepting 2 which have a different prior range: the surface emissivity k_{emis} has an interval between 0.94 and 1, since it is physically valid within these values and hum_{cste} , where the interval is wide enough to explore a more adapted value to enhance variable estimation.

Internal calculations of SECHIBA-YAO 1D use the true value of parameters; however, in the experiments presented in the next section, parameter posterior values are given in their normalized form.

Parameter	Description	Prior Value	Prior range	Unit
Internal Parameters				
hum_{cste}	Root extraction potential constant	4 (PFT 11)	0.2-16	m^{-1}
$rsol_{cste}$	Evaporation resistance	33000	16500-49500	s/m^2
dpu_{cste}	Total depth of soil water pool	2	1-3	m
min_{drain}	Diffusion between reservoirs	0,001	0,0005-0,0015	-
mx_{eau}	Maximum water content	150	75-225	Kg/m^3
Multiplying Factors				
k_{emis}	Surface Emissivity	1	0.94-1	-
k_{capa}	Soil Capacity	1	0.5-1.5	-
k_{cond}	Soil Conductivity	1	0.5-1.5	-
k_{rveg}	Vegetation Stomatal Resistance	1	0.5-1.5	-
k_{z0}	Roughness height	1	0.5-1.5	-
k_{albedo}	Surface albedo	1	0.5-1.5	-

Table 7. 1 SECHIBA parameters studied in this work. There are five internal parameters, involved in the model estimations and six multiplying factors that are imposed. In addition the prior range is specified for each parameter, defining the prescribed interval of variations.

7.3 LST data assimilation with parameter standard values

7.3.1. Simulated vs. Observed measurements

Brightness temperature

In SMOSREX site, a thermal infrared radiometer (TIR), functioning in the 8-14 μm spectral band, measures the surface brightness temperature (TB) at 60° pointing out south, in a fallow parcel (de Rosnay et al, 2006). Thus, we can expect directional effects in the measurements (François et al. 1997), especially during the periods of partial canopy coverage. In SECHIBA-YAO 1 D, the simulated LST is hemispheric and does not account for solar configuration and viewing angle effects, it is also integrated on all the solar spectrum. In this case, it does not correspond exactly to TB . However, there is a direct link between both variables. If we neglect the directional effects, the total energy emitted by the surface (Rad) can be computed using the following expression

$$Rad = k_{emis} \varepsilon \sigma LST^4 + (1 - \varepsilon k_{emis}) LW_{down} \quad (7.1)$$

In equation (7.1), ε is the surface emissivity, k_{emis} is the multiplicative factor for the emissivity and LW_{down} is the long-wave incident radiation, which is an input forcing of SECHIBA. Svendsen et al., (1990) proposed a transfer function to link the surface emitted radiance towards an observed brightness temperature measured in the limited [8-14] μm

spectral band. The empirical formulation is given by Eq. 7.2.

$$TB = \left(\frac{Rad - 7.84}{6.7975 \cdot 10^{11}} \right)^{0.2} \quad (7.2)$$

This formulation was added into SECHIBA-YAO1 D to allow the calculation of the brightness temperature and the comparison with the observed TB from SMOSREX. In order to do so, a module was added to SECHIBA-YAO 1D. This is done using YAO and adding a subgraph to the initial modular graph. As previously explained, this module contains the forward and backward expression (Jacobian) allowing the direct assimilation of TB , as presented in Fig.7.1.

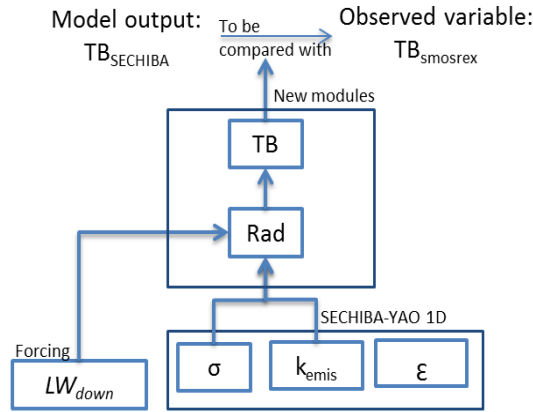


Figure 7. 1 Sub-graph added to the modular graph of SECHIBA-YAO 1D in order to simulate TB and compare it, to SMOSREX TB observations.

The first task before starting 4DVAR is to compare the behavior of our simulations with respect to the measurements. This gives us an idea about any possible corrections within the observations prior to the assimilation process, with the goal of improving its performance.

A comparison was done (Fig.7.2) using the data from 02/06/2006 to 09/06/2006. This particular sequence was chosen since the forcing was not affected during this week with any particular phenomenon (strong winds, clouds or heavy rain). The results show that SECHIBA-YAO 1D underestimates TB during the morning and overestimates it during the afternoon and night, compared to the SMOSREX observations, in other words, the diurnal amplitude is underestimated in SECHIBA compared to the measurements. Systematic differences may suggest directional effects in the observations linked to solar viewing effects which cannot be represented in our simple way to calculate LST .

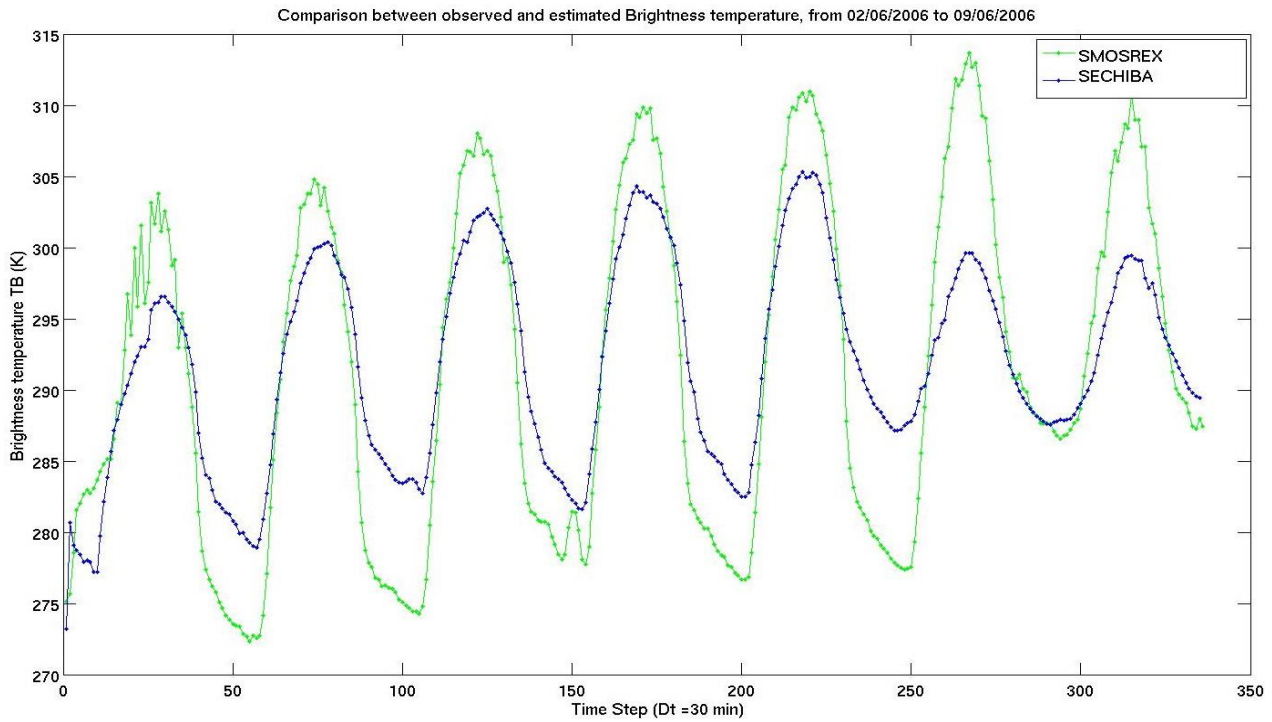


Figure 7. 2 Comparison between simulated (blue) and observed (green) *TB*, from 02/06/2006 to 09/06/2006. Abscissa corresponds to time steps (336 for a week) and ordinate corresponds to brightness temperature in Kelvin.

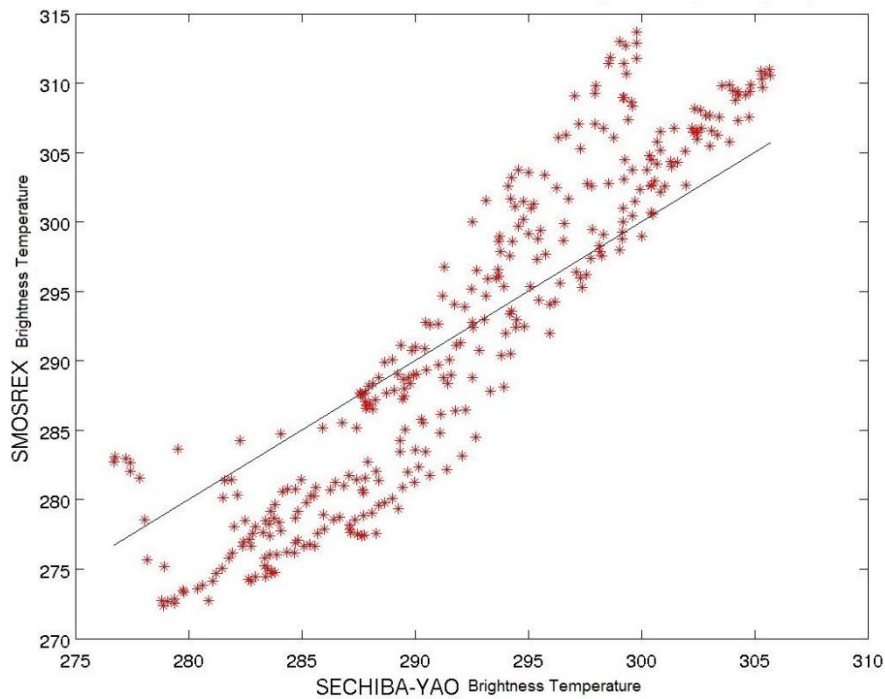


Figure 7. 3 Scatter plot of simulated vs. observed *TB*, for 2006. Abscissa corresponds to estimated *TB* and ordinate corresponds to SMOSREX measurements of *TB*, during 2006.

The scatter plot between the observed and simulated *TB* during 2006 is presented in Fig.7.3. The cluster is far from the identity function: the scatter plot shows a deviation from

the identity and a larger spread. In order to come up with these differences, a correction process called CDF-Matching is applied, with the aim of reducing the discrepancies between variables and observations. This method, proposed by Reichle et al., (2004) match the dynamics of a particular measurement. It modifies the observations making their statistical distribution closer to the one from the model. In their work, Reichle et al., (2004), applied it to soil moisture measurements. Another work from Parrens (2013) applied this filter technique to microwave observations, using SMOSREX measurements, with promising results. The same principle is implemented to our measurements of *TB*. CDF-Matching was applied to the year 2006 entirely, since a large number of realizations are needed (Scipal et al., 2008). A number of polynomials with several orders were computed with SMOSREX data. We choose finally a first order fit which gives an acceptable result. Observations are then rescaled in order to fit the computed polynomial. The resulting pseudo observations after CDF-Matching and the observed *TB* are shown in Fig.7.4.

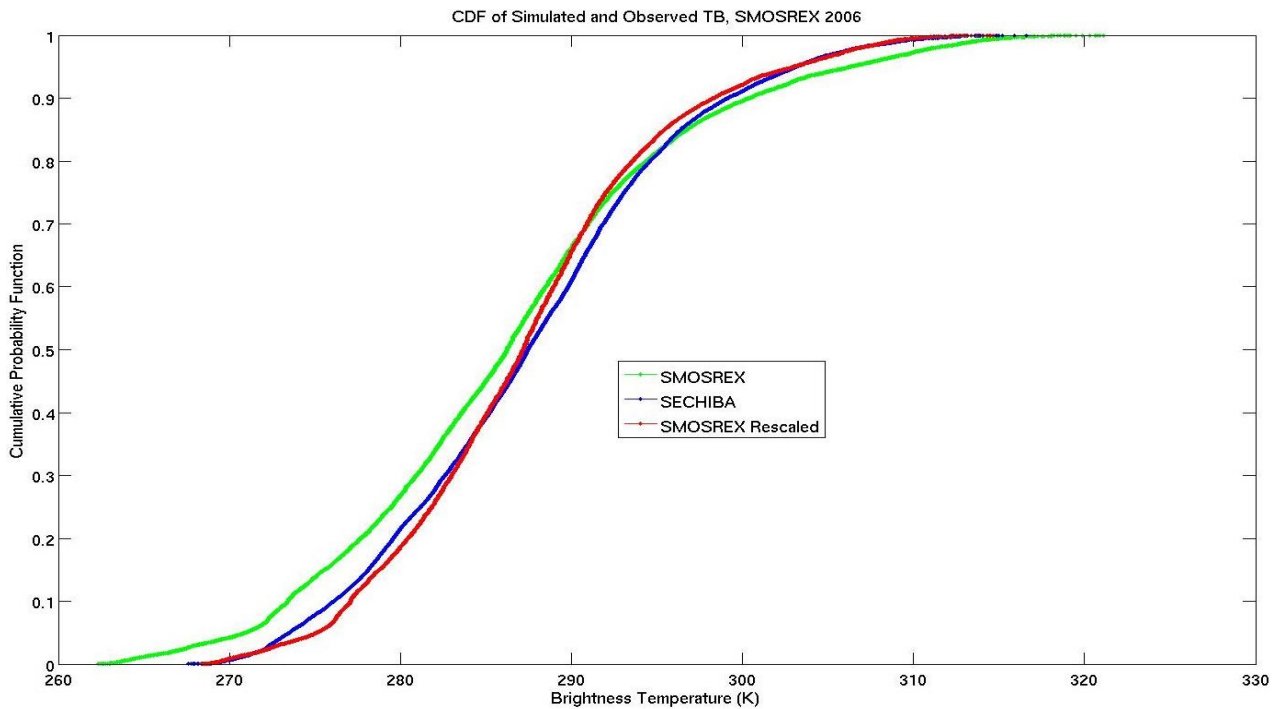


Figure 7. 4 Comparison of the Cumulative distribution function (CDF) of the brightness temperature (*TB*) estimated with SECHIBA (blue), observed in SMOSREX (green) and rescaled SMOSREX after CDF-Matching (red), for the year 2006.

Observations rescaled with CDF-Matching have lower diurnal amplitudes of variation. They fit better the distribution of SECHIBA estimation, as it can be seen in Fig.7.4 (for the year 2006) and Fig.7.5 (for a time series of *TB* taken during the week from 02/06/2006 to 09/06/2006). Nevertheless their accuracy decline showing the same drawbacks on the time period studied: underestimation at noon and overestimation at night.

We compute the RMSE as the root of the square difference between observations and estimations, divided by the number of observations. In addition standard deviation from SMOSREX prior and after CDF-Matching is computed. The prior RMSE, bias and standard deviation were 5.82 K, 14.74 K and 10.40 K. Posterior values of RMSE, bias and standard deviation were 3.28 K, 0.013 K and 8.19 K; showing that the gap between real measurements and estimations of TB is reduced. CDF Matching builds pseudo observations closer to the simulated variables (TB). In the following we will use these rescaled values of TB in our assimilation experiments.

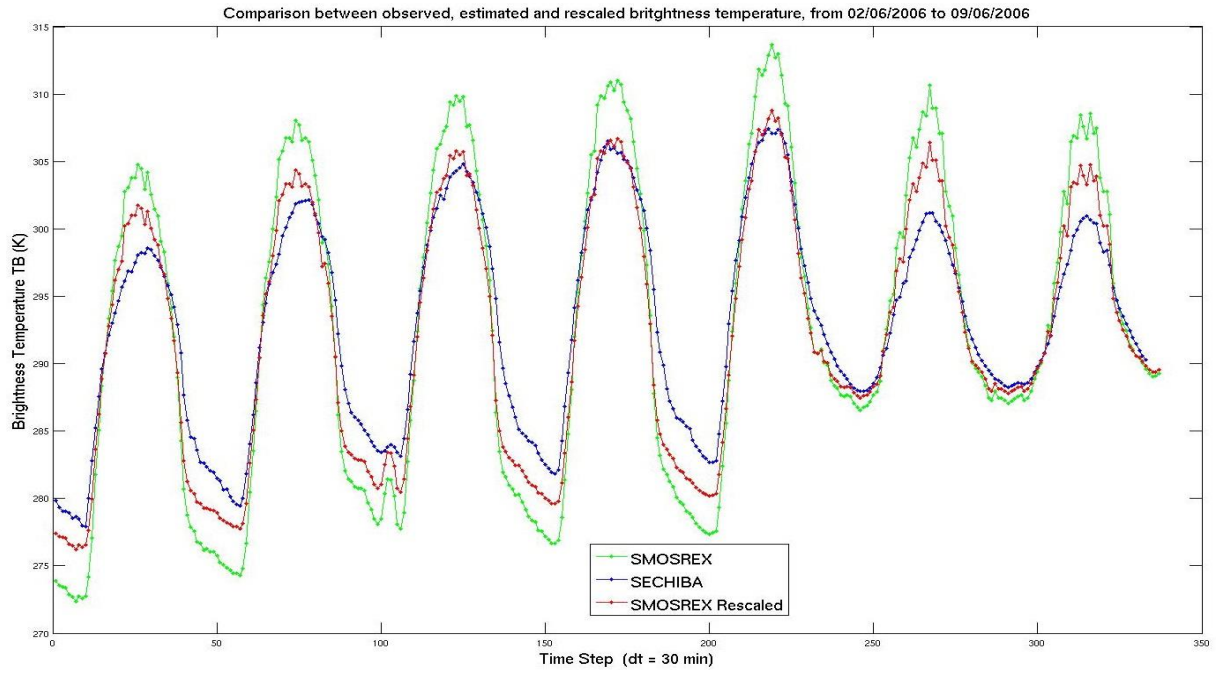


Figure 7. 5 Time series from 02/06/2006 to 09/06/2006 for TB estimated by SECHIBA (blue), observed in SMOSREX (green) and rescaled after CDF-Matching for SMOSREX measurements (red).

Soil moisture

Soil moisture (θ_s) profile in SMOSREX is characterized by measurements at different depths. Delta T probes sensors (de Rosnay et al, 2006) were installed in the ground at 0-5cm, 10 cm, 20 cm, 30 cm, 40 cm, 50 cm, 60 cm, 70 cm, 80 cm and 90 cm. In order to obtain a comparable variable with SECHIBA, a cumulative quantity of water in the soil profile was computed, as an integration of the measurements at different depths, with the following equation

$$\theta_s^{obs} = \sum_{i=0}^{10} z_i \theta_i \quad (7.3)$$

In equation (7.3), z_i and θ_i represent the thickness and the water content in the layer i . The sum integrates the different measurements taken from the ten layers. Soil moisture profile is therefore roughly represented and the main temporal and spatial characteristics of infiltration and uptake are captured.

Once the water content from SMOSREX was obtained, the corresponding variable in SECHIBA-YAO 1D was computed, in order to compare it with the observations. In SECHIBA the soil is divided into two buckets: the upper bucket (W_u) and the bottom bucket (W_l). At each time step, water content from each layer is computed with Eq.2.20. In order to obtain θ_s an integration of both layers divided by the maximum quantity of water allowed in the profile is computed, with the following expression

$$\theta_s = \frac{W_l + W_u}{300} \quad (7.4)$$

θ_s^{obs} and θ_s were normalized between 0 and 1 and thus are easily comparable. Both magnitudes represent the soil moisture, quantifying the moisture condition within the underlying soil profile. It must be kept in mind that this representation does not account for differences in absolute values of soil moisture –playing a role in thermal dynamics and H – since the maximum quantity of water varies with the parameters.

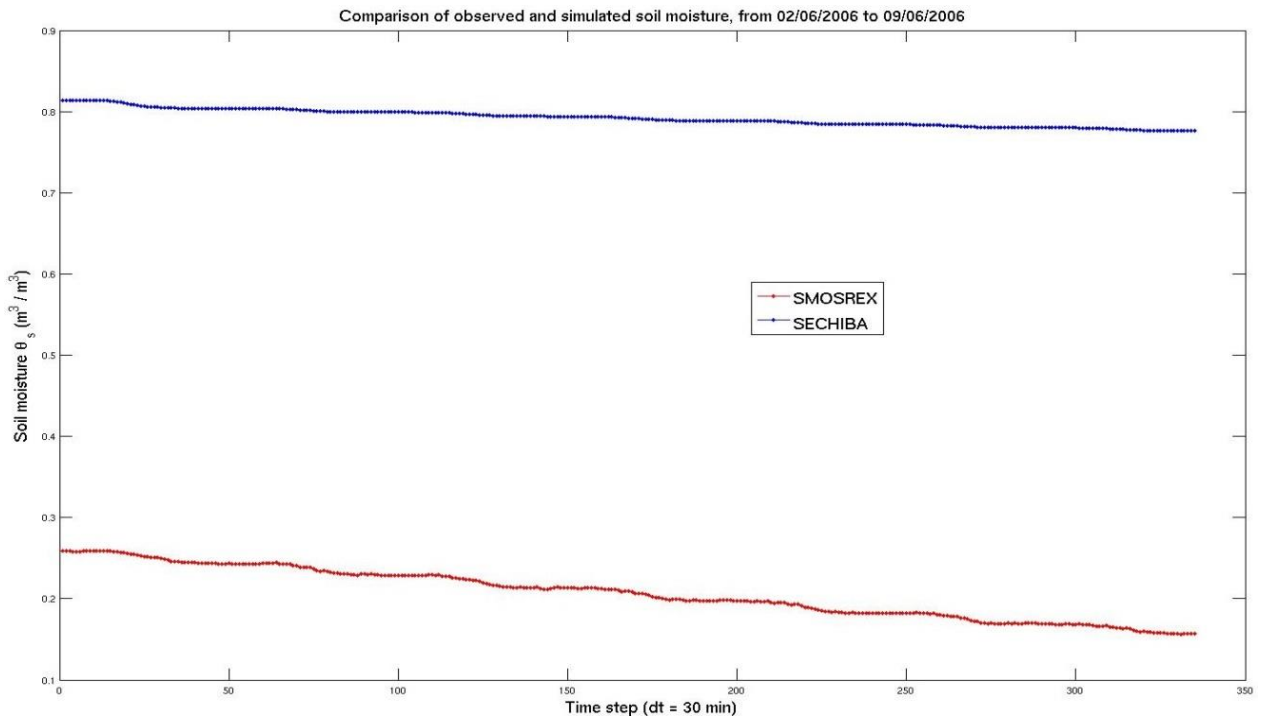


Figure 7. 6 Comparison between simulated (θ_s) and observed (θ_s^{obs}) soil moisture, from 02/06/2006 to 09/06/2006, using SMOSREX measurements.

We present in Fig.7.6 the estimation by SECHIBA-YAO 1D of θ_s compared to the soil moisture (θ_s^{obs}) observations during the week from 02/06/2006 to 09/06/2006. It can be seen that SECHIBA hydrology is not correctly simulated (blue curve): the soil moisture is too large and the daily variations too small, compared to θ_s^{obs} (red curve). Perhaps by adjusting parameters controlling the water stress, SECHIBA could evaporate or transpire the excess of stored water.

7.3.2. Brightness temperature sensitivity analysis

The gradient sensitivity analysis is applied using SMOSREX forcing in order to retrieve the most sensitive parameters to TB . As in Chapter 5, gradients are computed with respect to model parameters, using the adjoint model, during the 2nd June 2006. The idea is to build, as in Chapter 5, the parameter hierarchy of the most influential parameters to TB . In Fig.7.7, gradients of the brightness temperature with respect to model parameters are presented.

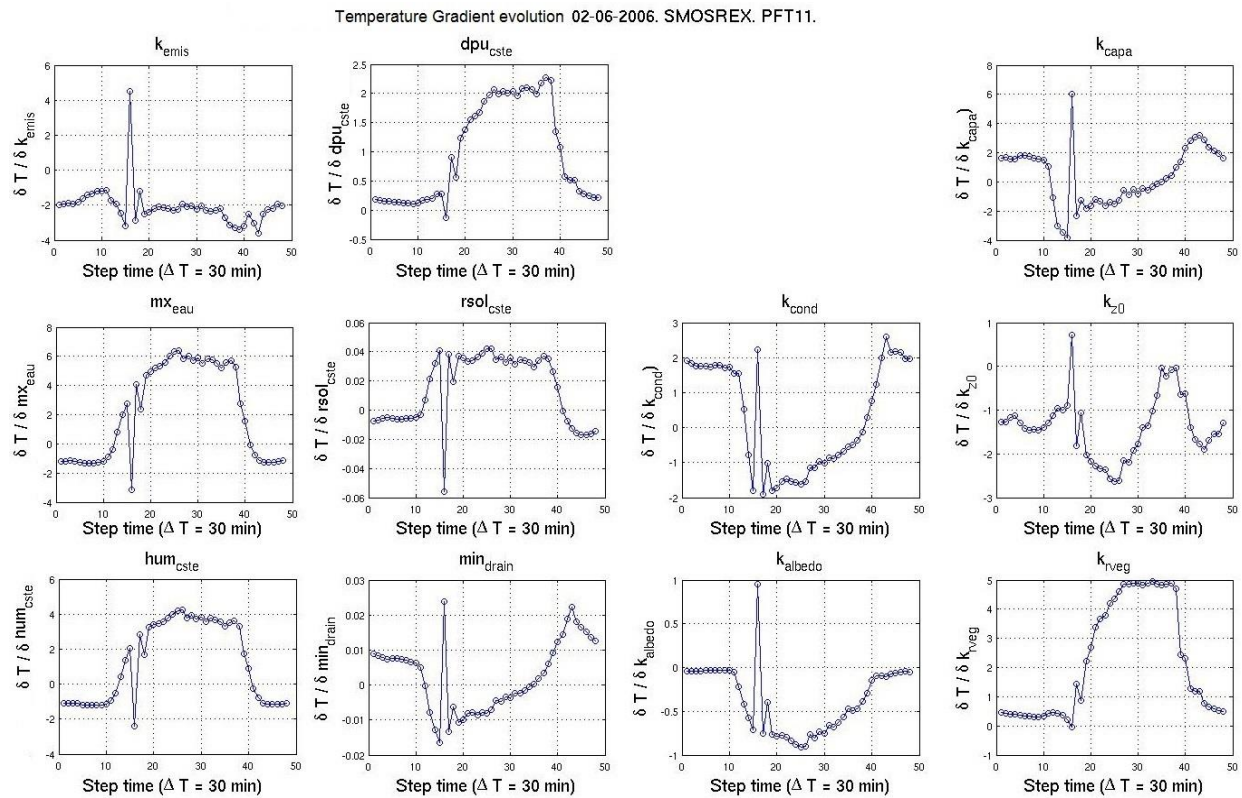


Figure 7. 7 Gradient sensitivity analysis using SMOSREX forcing, with PFT 11, the 02/06/2003. The sensitivities were computed on TB . Curves represent TB derivative with respect to each parameter given by the adjoint model, every half hour over a day.

In Table 7.2 the 11 parameters ranked in order of influence are presented, according to the gradients magnitude, computed in Fig.7.7. Parameters hierarchy revealed that the highest

gradient values using SMOSREX site matches the gradient values results with FLUXNET sites, however the order in the hierarchy of the most sensitive parameters is not the same. Even if the magnitude of gradients may vary, parameters related to the thermodynamic are among the most sensitive to brightness temperature for both sites (k_{emis} , k_{cond} , k_{capa}).

Variable	Hierarchy
Brightness temperature	mx_{eau} , hum_{cste} , k_{capa} , k_{emis} , k_{cond} , k_{rveg} , dpu_{cste} , k_{z0} , k_{albedo} , $rsol_{cste}$, min_{drain}

Table 7. 2. Parameter hierarchy using SMOSREX forcing, for brightness temperature.

The least sensitive parameter (min_{drain}), is not optimized in the following experiments, only the ten most sensitive parameters are taken into account.

7.3.3. Brightness temperature assimilation during a single day

Experiment Definition

This experiment consists in the realization of successive assimilation during a month. The assimilations are performed in a time window of one day. At the end of the assimilation, the window is moved to the next day, the system is reinitialized and next assimilation is run. The initial condition of the model state variables is different from one day to another. By controlling the most sensitive parameters and by using rescaled observations of TB , the parameter variability is obtained as well as the error evolution through a month.

Experiment	Observation	Parameters
1	Rescaled TB , every 30 minutes	mx_{eau} , hum_{cste} , k_{capa} , k_{emis} , k_{cond} , k_{rveg}

Table 7.3 Configuration for Experiment 1. The six most sensitive parameters in the calculation of TB are optimized. The parameter values are set to their prior values (Table 7.1)

In Experiment 1, as shown in Table 7.3, the assimilation of the rescaled TB is performed with a sampling of 30 minutes (every time step) is performed. The six most sensitive parameters to TB are controlled, based on the parameter hierarchy computed in the previous section. These parameters are mx_{eau} , hum_{cste} , k_{capa} , k_{emis} , k_{cond} and k_{rveg} . SMOSREX measurements were taken from 02/06/2006 during a month, with PFT 11.

Results

For each assimilation test, the prior and posterior RMSE calculated between simulated and observed measurements are computed. Fig.7.8 shows their values through a month. There are days where no observations are available, thus no RMSE is computed. From Fig.7.8, it can

be observed that TB estimations are improved with the one-day assimilation scheme. Regarding the other variables, R_n and θ_s have no variation, H is improved every day excepting the 10th day and LE is improved excepting the 20th and the 22th day.

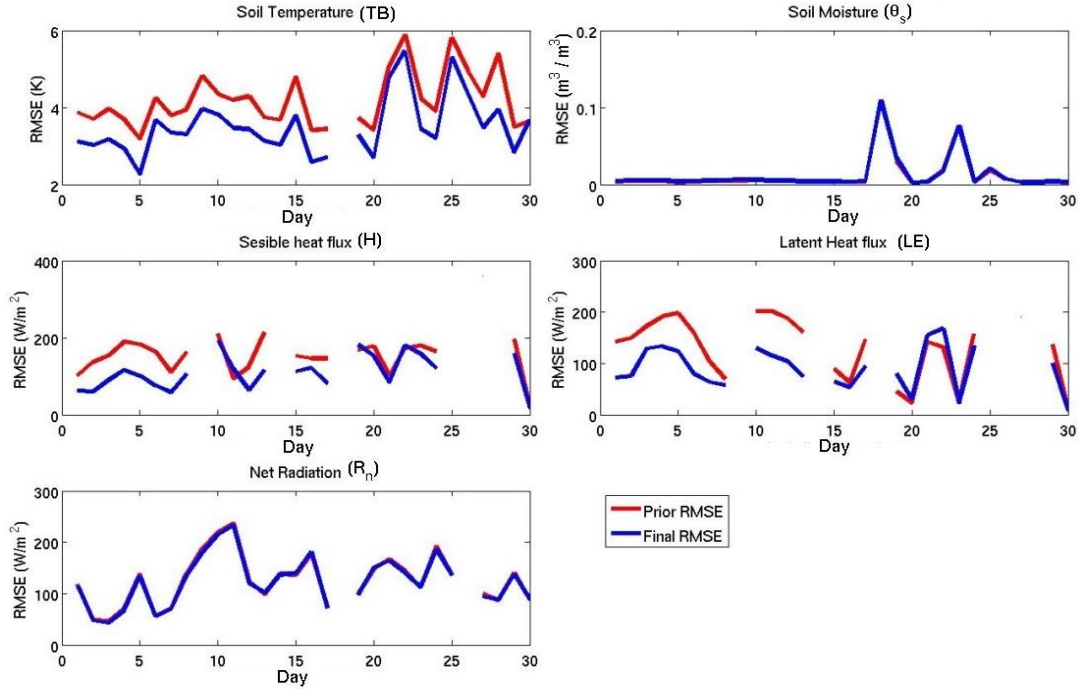


Figure 7. 8. Prior and posterior assimilation RMSE values for model variables. Abscissa corresponds to the day of the assimilation, beginning from the 02/06/2006 to 02/07/2006. Ordinate corresponds to RMSE values. Red curves stand for the RMSE of the different variables prior to assimilation. Blue curves are the RMSE after assimilation.

Nonetheless, these improvements are obtained assigning parameters values at the edge of their prior range. Fig.7.9 shows the cumulative distribution of the parameters in the interval. Parameters k_{capa} , mx_{eau} and k_{cond} are concentrated around the lower boundary of the interval (0.5) in most of the results. Same results are obtained for k_{rveg} , whose posterior values are most of the time in the upper boundary of the interval (1.5). Although k_{emis} prior range interval is restrained between 0.94 and 1, their values remain more or less stable.

In summary, Experiment 1 allows us to retrieve the evolution during one month of the parameters posterior values in successive assimilations of TB . This is important since it allows us to measure the limits of the assimilation process when taking a one-day assimilation time window, in order to reduce the gap between estimated and rescaled TB observations. A general improvement in simulated variables was obtained at the expense of parameter values in the boundaries of the prior range.

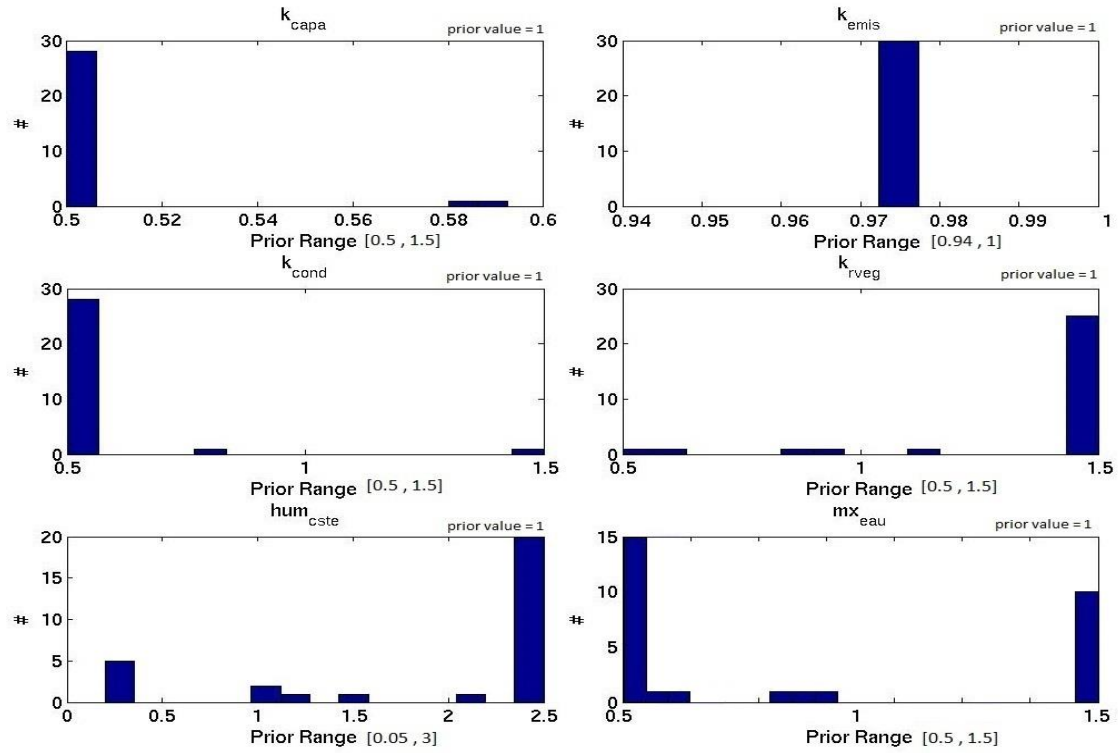


Figure 7. 9 Parameters posterior values, when assimilating brightness temperature. Plots are the cumulative histograms of the parameter posterior values obtained after one day assimilation.

7.3.4. Brightness temperature assimilation during a week

The results of the previous section may suggest that a one-day assimilation time window is not enough, in order to calibrate properly control parameters, since improvements in variable estimation were obtained at the expense of assigning parameter values at the edge of the parameters interval.

Experiment Definition

The different scenarios for the assimilation experiments with a seven-day window are presented in Table 7.4. As in Experiment 1, SMOSREX forcing was used with vegetation type grassland (PFT 11), from 02/06/2006 to 09/06/2006, with an observation frequency of 30 minutes (every time step). In Experiment 2, the six most sensitive parameters to TB are controlled (section 7.3.2). Experiment 3 optimize the ten most sensitive parameters.

Experiment	Observation	Parameters
2	Rescaled TB	mx_{eau} , hum_{cste} , k_{capa} , k_{emis} , k_{cond} , k_{rveg}
3	Rescaled TB	k_{capa} , k_{emis} , k_{cond} , k_{rveg} , hum_{cste} , mx_{eau} , k_{z0} , k_{albedo} , dpu_{cste} , $rsol_{cste}$

Table 7. 4. Scenarios for Experiments 2 and 3. Both experiment used rescaled TB observations, with a 30 minutes sampling.

Results

The RMSE computed at the end of the assimilated process is presented in Table 7.5, the posterior parameter values retrieved after each assimilation experiment are presented in Table 7.6. Posterior RMSE of TB in experiment 2, when only six parameters are controlled, is 2.13 K, which is slightly better than the result obtained in experiment 3, where the posterior RMSE is equal to 2.33 K. By observing the posterior RMSE for θ_s , R_n , LE and H , both experiments degrade the prior estimations, except for R_n in experiment 2 (RMSE = 64.46 W/m^2). By controlling less parameters, it seems that the assimilation process degrades less the different variables. The amplitude of the RMSE is considerably high for each of the diagnostic variables.

Experiments		RMSE				
		TB(K)	θ_s (m^3/m^3)	R_n (W/m^2)	LE (W/m^2)	H (W/m^2)
2	Prior	2.87	$5.42 \cdot 10^{-1}$	66.35	54.69	68.28
	Posterior	2.13	$6.51 \cdot 10^{-1}$	64.46	69.81	92.14
3	Posterior	2.33	$6.52 \cdot 10^{-1}$	66.52	71.84	93.59

Table 7. 5. Prior and posterior RMSE for Experiments 2 and 3. Gray cells in the table show the best RMSE values for each scenario.

Table 7.6 shows that most of the retrieved parameters are at the boundaries of the interval, suggesting that the improvements obtained in the estimation of TB are obtained at the expense of assigning parameters values at the edge of the interval.

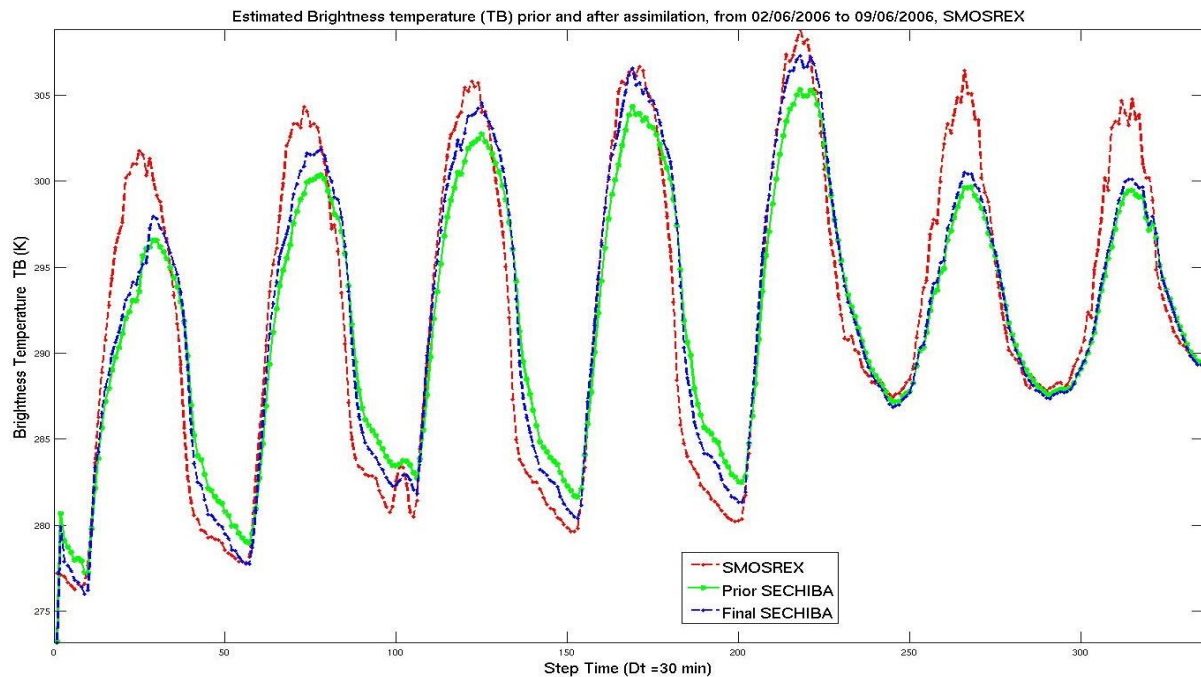


Figure 7. 10. Time series for TB before and after assimilation. Results correspond to Experiment 2. Red curves are the rescaled SMOSREX measurements. Green curves are SECHIBA estimations with the parameters prior values. Blue curves are the estimated TB with the optimized parameters.

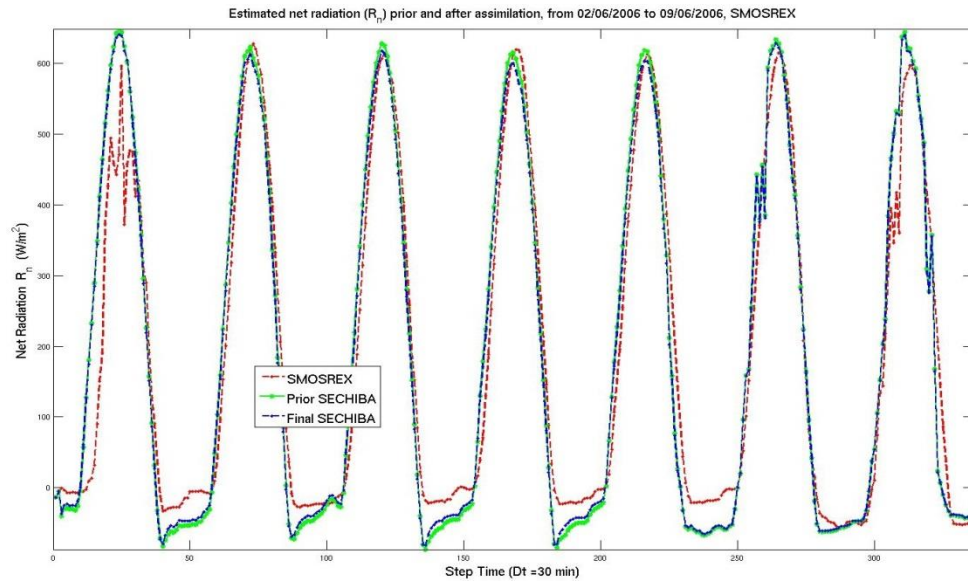
Experiments	k_{emis}	k_{veg}	k_{cond}	k_{capa}	k_{z0}	k_{albedo}	mx_{eau}	hum_{cste}	dpu_{cste}	$rsol_{cste}$
2	1	1.5	0.5	0.5	-	-	1.5	1.1	-	-
3	1	1.5	0.5	0.5	0.5	0.81	1.5	1.1	1.5	0.99

Table 7. 6. Parameter posterior values for Experiments 2 and 3, after assimilating rescaled TB . Prior values for each parameter are normalized to 1. Gray cells show parameter at the boundaries.

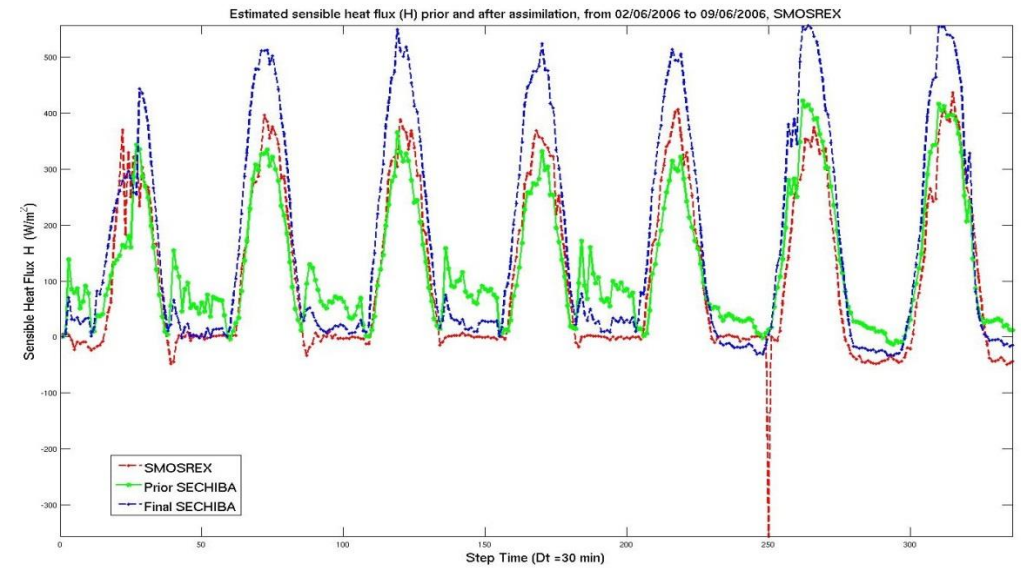
We use the parameters final value of Experiment 2 (when the reduction of TB RMSE is most significant), and perform comparisons between the estimated variables vs. SMOSREX observations prior and after the assimilation process. It can be observed that:

- Estimated TB after assimilation is closer to the SMOSREX measurements (Figure 7.10)
- R_n , shows almost no change in the estimation of the flux prior and posterior the assimilation. SECHIBA prior estimation underestimates this flux (Figure 7.11.a).
- For the rest of the variables (H , LE and θ_s) prior estimations are degraded (Figures 7.11.b, 7.11c, 7.11d). H is overestimated systematically, similar to θ_s , which is degraded because of LE , which is underestimated during the assimilation period.

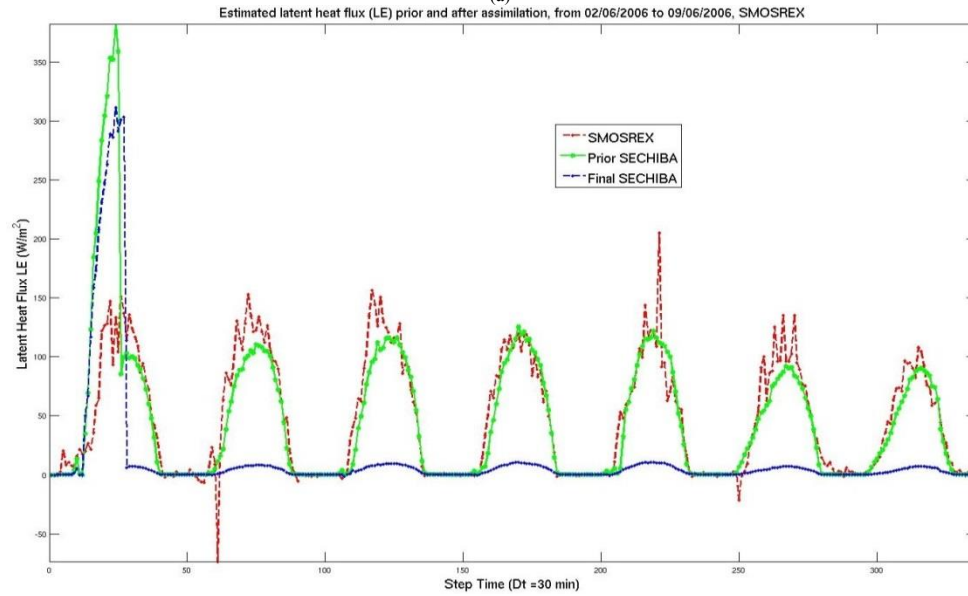
The results show how the final parameter values are not suited to improve all variables.



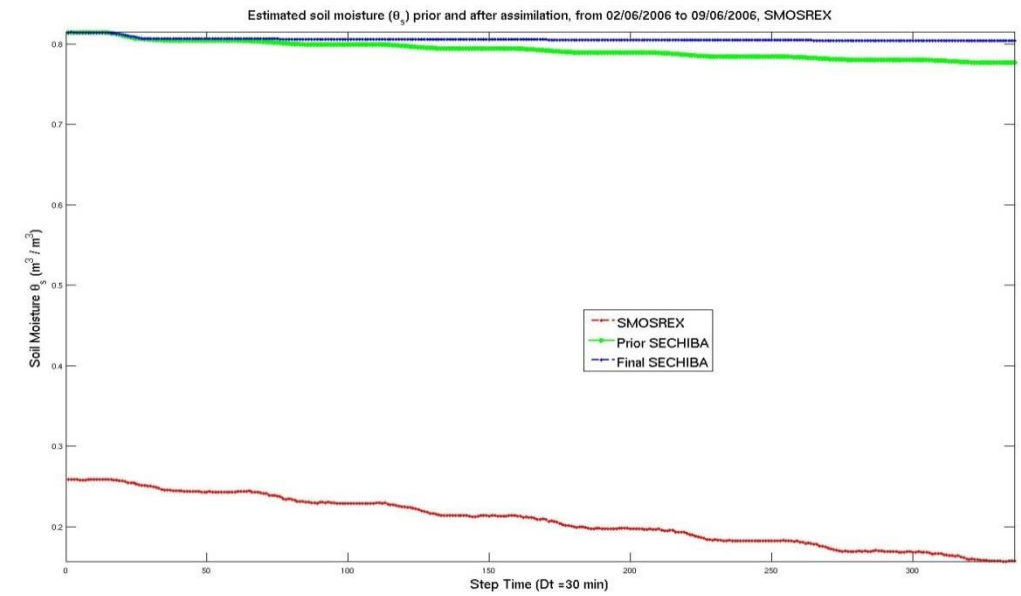
(a)



(b)



(c)



(d)

Figure 7. 11. Time series for R_n (a) H (b), LE (c) and θ_s (d) before and after assimilation, for Experiment 2, from 02/06/2006 to 09/06/2006. Red curves are the SMOSREX measurements. Green curves are estimations with the parameters prior values. Blue curves are the variables with the optimized parameters.

7.3.5. Discussion

Inspecting the prior values of simulated TB with SECHIBA shows that this variable is underestimated compared to rescaled SMOSREX measurements. In an assimilation experiment, the increase of TB is obtained by decreasing the two parameters k_{capa} and k_{cond} and by increasing k_{rveg} and hum_{cste} . These parameters values lead to decrease transpiration and increase sensible heat flux, which contribute to stress the model errors concerning these fluxes, as well as θ_s , since prior values are overestimated. This is exactly what it is obtained from Experiments 2 and 3: doing so the TB is improved while LE , H and θ_s are systematically degraded.

To better assess the functioning of the assimilation tool and to analyze if the contribution of TB assimilation would be altered with different initial conditions, a separate region of the parameter space was tested. As the uncertainty of hum_{cste} is high compared to the rest of the model parameters (its value is merely empirical) we choose to modify only this parameter that impacts the evapotranspiration.

7.4 LST variational data assimilation with different prior values

In the following assimilation experiments, a prior value of hum_{cste} of 0.4 is taken instead of $hum_{cste} = 4$. This value implies a larger root extraction potential and larger transpiration, allowing us to analyze the impact of the TB assimilation in such conditions. In the next section, comparison between estimation and observations for TB and θ_s are presented.

7.4.1. Simulated vs. Observed measurements

Different simulated vs. observed variables are presented, in order to better assess the impact of hum_{cste} in the model simulations. For the different variables, two sets of simulations are presented: the first one using the prior value of all parameters presented in Table 7.1 and the second by changing only hum_{cste} to 0.4.

Brightness temperature

The Fig.7.12 presents a comparison between original TB observations of SMOSREX and estimations with SECHIBA using hum_{cste} equal 4 and 0.4. It is observed that SECHIBA underestimates TB with $hum_{cste} = 4$ (blue curve), by comparing it to the observed TB in SMOSREX (green curve). Degradation is more pronounced in TB estimations with $hum_{cste} = 0.4$ (magenta curve) than the estimation with the standard value, and the same systematic

discrepancies are observed.

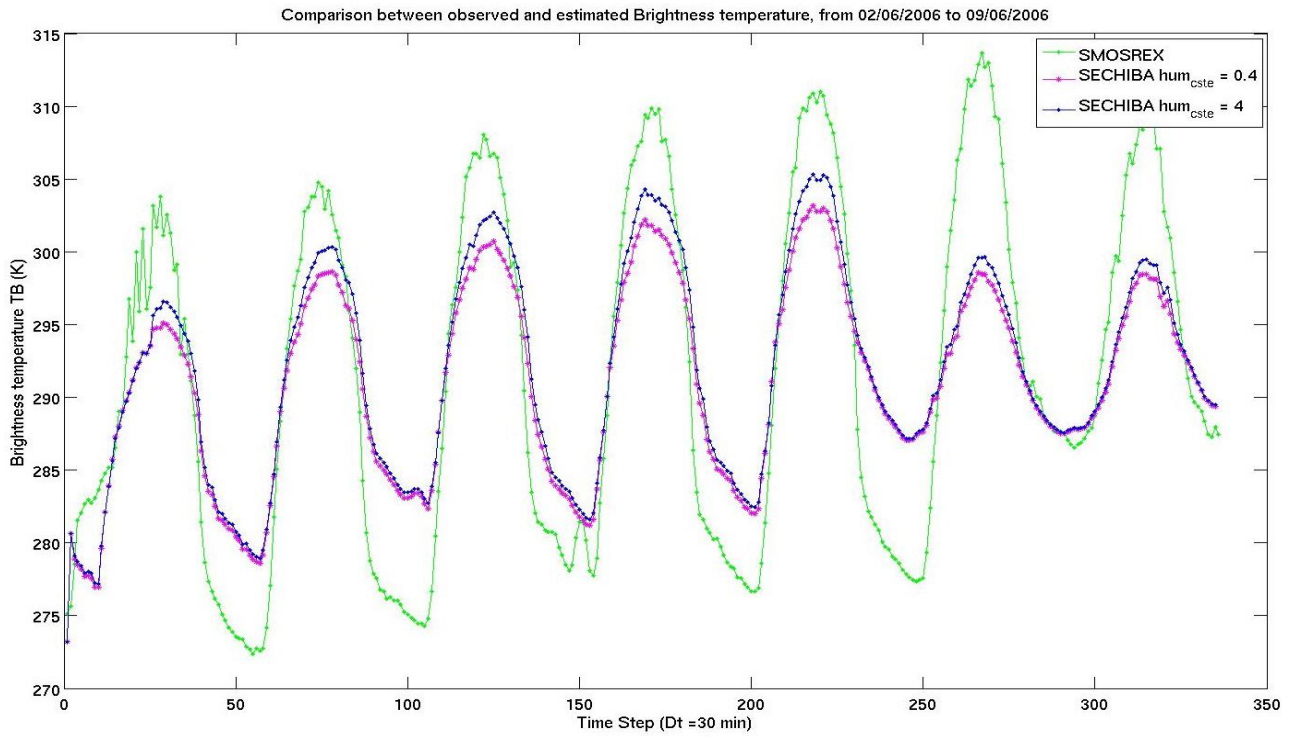


Figure 7. 12. Comparison between SECHIBA TB with $hum_{cste}=0.4$ (magenta), SECHIBA TB with $hum_{cste}=4$ (blue) and SMOSREX original measurements of TB (green), from 02/06/2006 to 09/06/2006, using SMOSREX forcing. Abscissa corresponds to time steps (336 for a week) and ordinate corresponds to brightness temperature in Kelvin.

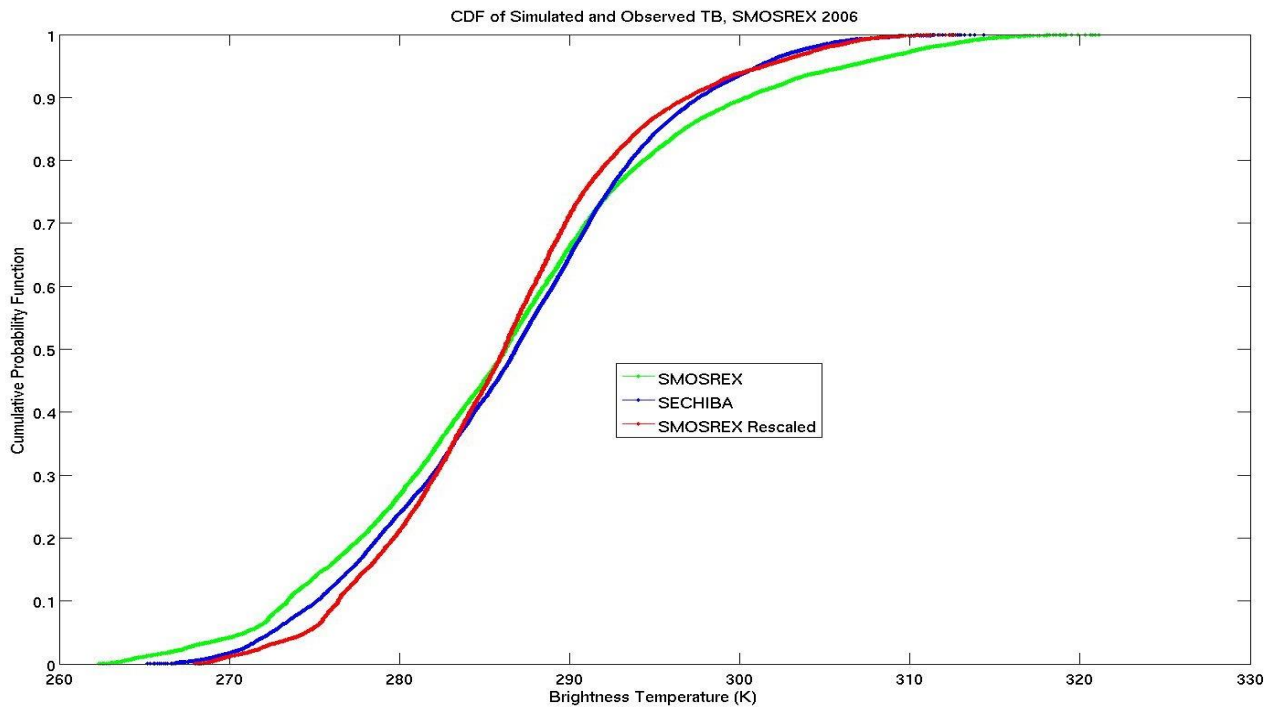


Figure 7. 13. Cumulative distribution function (CDF) of the brightness temperature (TB) estimated with SECHIBA (blue), SMOSREX (green) and rescaled SMOSREX after CDF-Matching (red).

In order to come up with these differences, CDF-Matching is applied, as described in section 7.3, using TB estimations with $hum_{cste} = 0.4$. Observations are then rescaled in order to fit the computed polynomial. CDF-Matching was applied to the whole of year 2006, using a first-order polynomial. The resulting CDF of estimations and observed TB after CDF-Matching are shown in Fig.7.13. Prior RMSE, bias and variance are 8.55 K, 8.75 K and 10.40 respectively. Posterior RMSE, bias and variance are 4.45 K, 0.009 K and 7.91 K. Observations rescaled with CDF-Matching have lower amplitudes of variation but they fit better the distribution of our estimation. When looking at extreme values, rescaled and simulated CDF are closer than the original observations vs. the simulations.

Soil moisture

The comparison between the observed and simulated θ_s , for hum_{cste} equal 0.4 and 4 is presented in Fig.7.14. Estimations with SECHIBA standard values (green curve) overestimate θ_s . Even though estimations with $hum_{cste} = 0.4$ (blue curve) are also overestimated, the daily variables fit better the SMOSREX observation (red curve) with the same dynamics as the SMOSREX measurements.

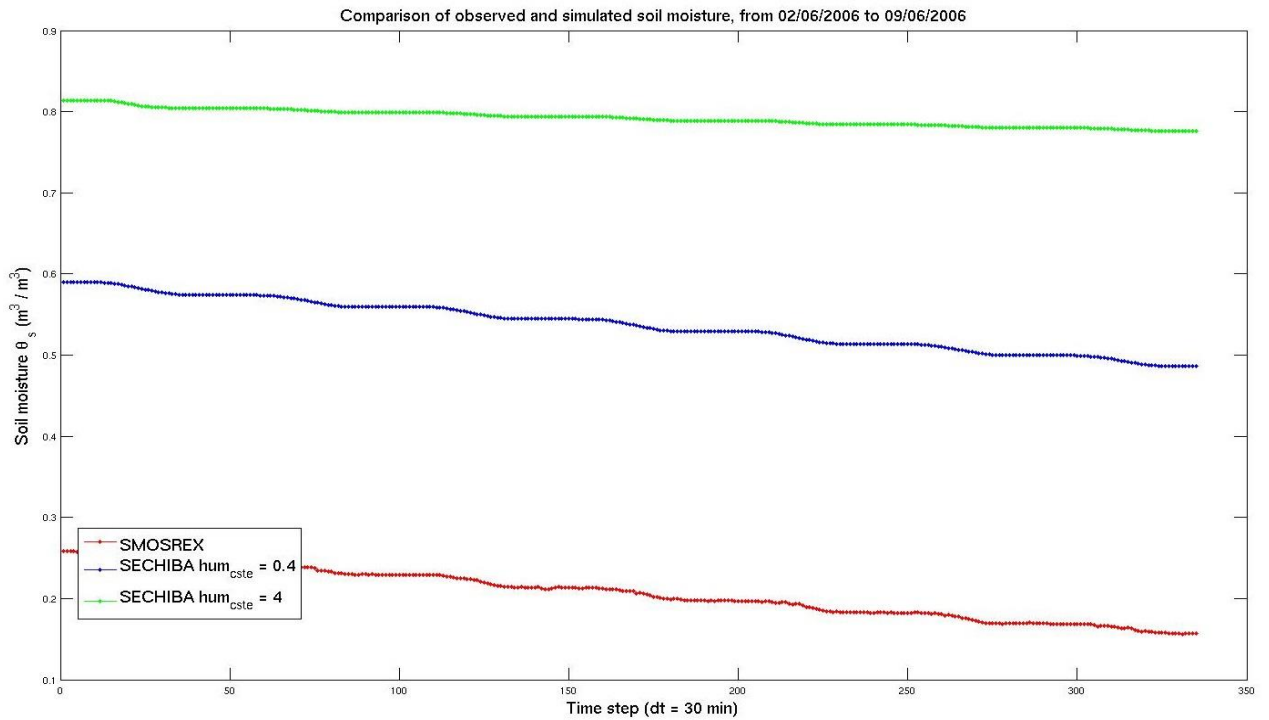


Figure 7. 14. Comparison between simulated (θ_s) and observed soil moisture (θ_s^{obs}), from 02/06/2006 to 09/06/2006, using SMOSREX measurements.

By reducing hum_{cste} , the potential root extraction is increased, leading to an increase in transpiration, causing a decrease in the soil water content. It is clear that in the scenario of

$hum_{cste} = 0.4$, the prior value is almost twice better, improving the performance of our model.

7.4.2. Brightness temperature sensitivity analysis

A gradient sensitivity analysis was performed using SMOSREX forcing in order to retrieve the most sensitive parameters to TB , with $hum_{cste} = 0.4$. As in Section 7.3.2, gradients are computed with respect to model parameters, using the adjoint model, during the 2nd June 2006. In Fig.7.15, gradients of the brightness temperature with respect to model parameters are presented.

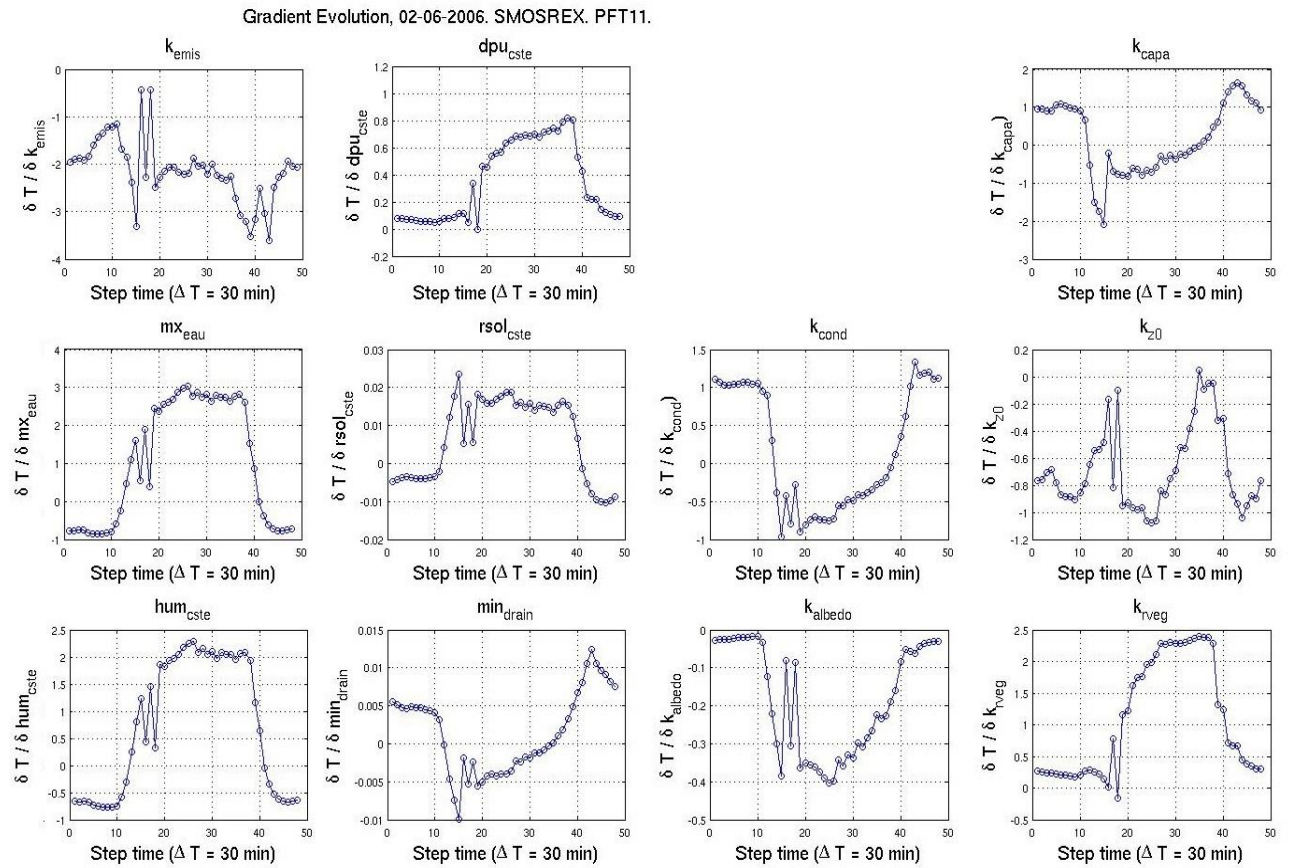


Figure 7. 15. Gradient sensitivity analysis using SMOSREX forcing, with PFT 11, the 02/06/2003. The sensitivities were computed on the TB , using $hum_{cste} = 0.4$. Curves represent TB derivatives with respect to each parameter given by the adjoint model.

In Table 7.7, the eleven parameters ranked in order of influence are presented, according to the gradients magnitude, as shown in Fig.7.15. Parameters hierarchy using $hum_{cste} = 0.4$ is the same as obtained in Section 7.3.2, where the standard prior value for hum_{cste} was used. This revealed that, although a different point in the parameter space is taken, the hierarchy remains the same even though magnitudes of the sensitivity are smaller in general.

Variable	Hierarchy
Brightness	$mx_{eau}, hum_{cste}, k_{capa}, k_{emis}, k_{cond}, k_{rveg},$
temperature	$dpu_{cste}, k_{z0}, k_{albedo}, rsol_{cste}, min_{drain}$

Table 7. 7. Parameter hierarchy using SMOSREX forcing, for brightness temperature

7.4.3. Brightness temperature assimilation during a single day

Experiment Definition

This experiment consists in the realization of successive assimilation during a month, using a time window of one day; similar to Experiment 1 presented in Section 7.3.3.

Experiment	Observation	Parameters
4	Rescaled TB, every 30 minutes,	$mx_{eau}, hum_{cste}, k_{capa}, k_{emis}, k_{cond}, k_{rveg}$ $hum_{cste} = 0.4$

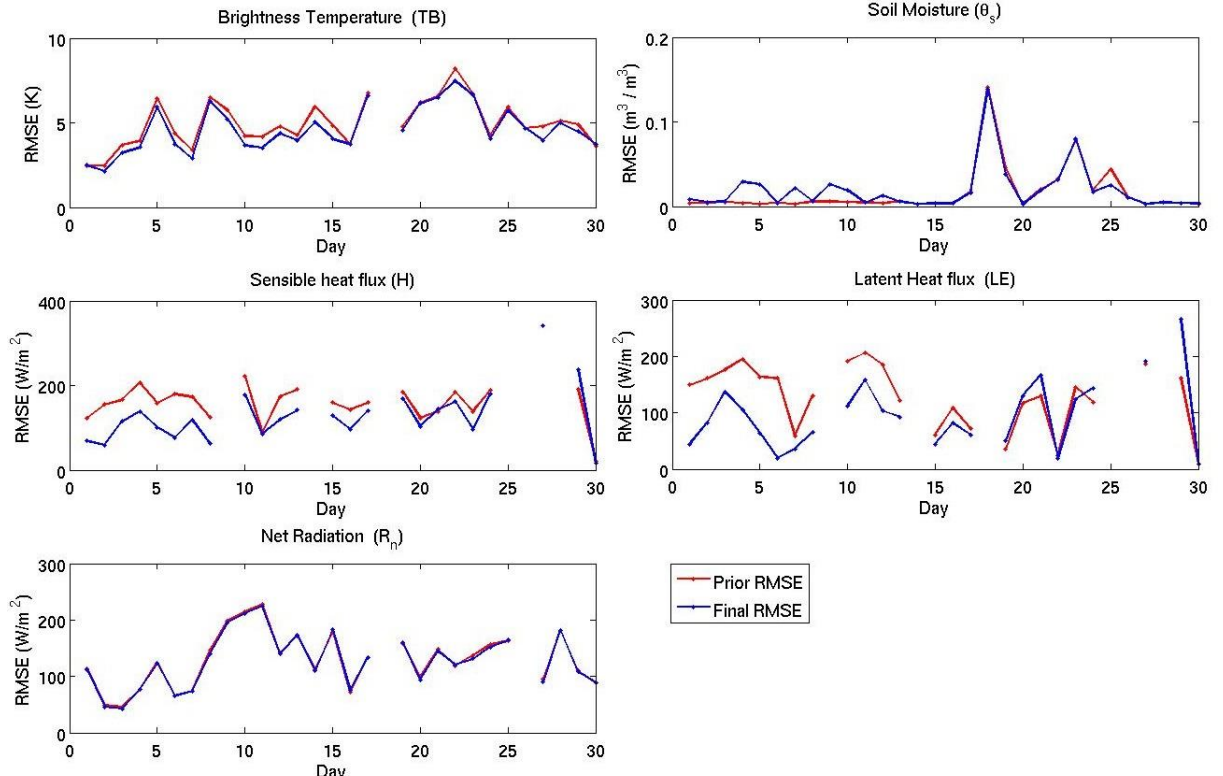
Table 7. 8. Experiment 4. The six most sensitive parameters of TB are optimized. The parameter values are set to their prior values, excepting hum_{cste} , which is equal to 0.4.

In Experiment 4, the six most sensitive parameters to TB are controlled, based on the parameter hierarchy computed in the previous section. Rescaled SMOSREX TB observations were taken from 02/06/2006 during a month, with PFT 11. Standard values for the different parameters are taken, excepting hum_{cste} , which prior value was equal to 0.4. This experiment uses the observations of TB rescaled, obtained after the CDF-Matching process.

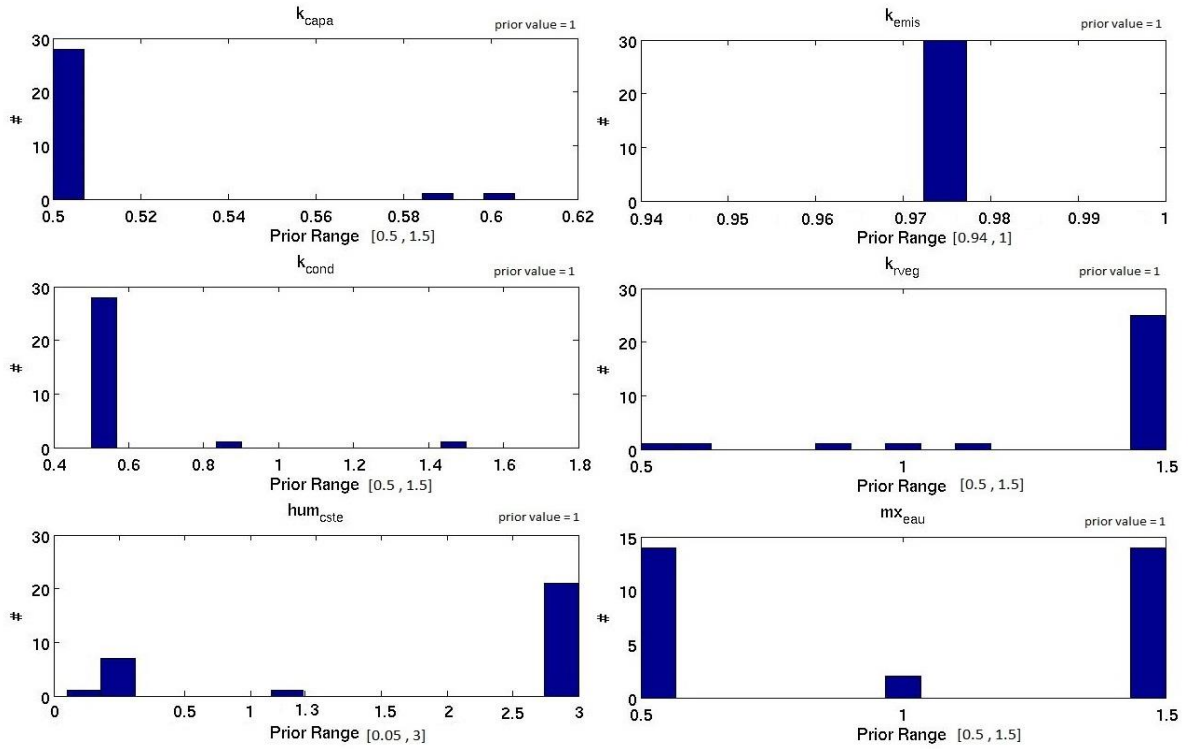
Results

As in Fig.7.8 we present for each day and each variable (H , LE , TB , R_n and θ_s) the RMSE computed using the rescaled SMOSREX TB measurements and the corresponding values estimated by SECHIBA at the end of the assimilation process (Fig.7.16.a). The posterior RMSE for H , LE and TB is reduced. R_n presents no variation. The θ_s values are degraded throughout the month. Fig.7.16.b shows the cumulative distribution of the parameters in the interval. The distribution of the parameter posterior values is similar to Experiment 1, where $hum_{cste} = 4$.

In summary, experiment 4 shows that a one-day assimilation of TB globally improves estimation of TB , H and LE . Parameter posterior values reach their boundary. In order to test the role of the assimilation windows, assimilation during a week of the most sensitive parameters is performed in Section 7.4.4 using $hum_{cste} = 0.4$.



(a)



(b)

Figure 7. 16. (a) Comparison of RMSE values prior and after assimilation, during a month. Red curves stand for the RMSE of variables prior to assimilation. Blue curves are the RMSE after assimilation. (b) Parameters posterior values, when assimilating *TB*. Plots are the cumulative histograms of the parameter posterior values obtained after one day assimilation.

7.4.4. Brightness temperature assimilation during a week

Experiment Definition

The different scenarios for the assimilation experiments with a seven-day window are presented in Table 7.9. As previous experiments, SMOSREX forcing was used with a grassland (PFT 11) vegetation type, from 02/06/2006 to 09/06/2006, with an observation frequency of 30 minutes (every time step). In Experiments 5, the six most sensitive parameters to TB are controlled. Experiment 6 controls the ten most sensitive parameters to TB .

Experiment	Observation	Parameters	Prior value
5	Rescaled TB	mx_{eau} , hum_{cste} , k_{capa} , k_{emis} , k_{cond} , k_{rveg}	$hum_{cste} = 0.4$
6	Rescaled TB	k_{capa} , k_{emis} , k_{cond} , k_{rveg} , hum_{cste} , mx_{eau} , k_{z0} , k_{albedo} , dpu_{cste} , $rsol_{cste}$	$hum_{cste} = 0.4$

Table 7. 9. Scenarios for experiments 5 and 6. Every experiment used rescaled TB observations, with a 30 minutes sampling. hum_{cste} is equal to 0.4.

In Experiments 5 and 6, prior value for hum_{cste} is 0.4. By comparing Experiments 5 with 6, the impact of controlling a larger parameter set can be measured. Moreover, by matching Experiments 2 and 3 with Experiments 5 and 6, the influence of changing the prior value of the parameter set can be accounted.

Results

Assimilation performances with rescaled TB observations during a week are presented in Table 7.5. In addition, the posterior values of parameters retrieved after each assimilation are presented in Table 7.6. Final RMSE of TB in Experiment 5, for which only six parameters are controlled, is 2.06 K. A slight improvement is obtained with ten controlled parameters (Exp. 6). However, the differences are marginal: one experiment does not outperform the others. Experiments 5 and 6 show a slight degradation of posterior θ_s RMSE. Improvements are obtained for the rest of the variables. RMSE final values for Experiment 6 are larger than those obtained in Experiment 5.

Experiments	Prior	RMSE				
		TB (K)	θ_s (m ³ /m ³)	Rn (W/m ²)	LE (W/m ²)	H (W/m ²)
5	Posterior	2.06	5.31.10 ⁻¹	64.45	50.7	55.79
6	Posterior	2.32	5.51.10 ⁻¹	67.01	71.5	92.3

Table 7. 10. Assimilation performances when assimilating rescaled using $hum_{cste} = 0.4$. By comparing both experiments, gray cells in the table show the best RMSE value for each scenario.

Experiment	k_{emis}	k_{veg}	k_{cond}	k_{capa}	k_{z0}	k_{albedo}	mx_{eau}	hum_{cste}	dpu_{cste}	$rsol_{cste}$
5	1	1.5	0.5	0.5	-	-	1.499	1.1	-	-
6	1	1.499	0.5	0.5	0.5	0.5	1.5	1.1	1.5	0.99

Table 7. 11. Parameter posterior values for experiments 5 and 6, after assimilating rescaled TB with $hum_{cste} = 0.4$. Prior values for each parameter are normalized to 1. Gray cells show parameter values at the limit of their prior range.

Parameter posterior values are presented in Table 7.11 for Experiments 5 and 6. It can be observed that most of the time the posterior values reach their boundaries. These results are similar to those in Experiments 2 and 3.

7.4.5 Discussion

In this section, the use of a different prior value for hum_{cste} was tested in several assimilations, with two different assimilation time windows. For both windows (one and seven days) posterior RMSE for TB , H and LE was reduced, when constraining control parameter with rescaled TB observations. Comparing Experiments 5 and 6, variables posterior RMSE all decreased except for a slight degradation of θ_s . Best performances were obtained by controlling the six most sensitive parameters (exp. 5). However, even if a higher reduction of variables posterior RMSE is obtained using hum_{cste} prior value of 0.4, some posterior parameters values reach the limit of their prescribed boundaries.

7.5. Analysis of the assimilation system through twin experiments

Experiments 1 to 6 show how TB assimilation alone brings a weak constrain on control parameters. Forcing the parameters to their prescribed limits increases variable estimation accuracy, but the minimum of the cost function is not reached.

Fig.7.17 shows the sampling of the cost function in the assimilation experiments with the standard values of the control parameters. We sampled the cost function by varying the parameters from $\pm 50\%$ of their initial value (here the initial value is always 1), from the period 02/06/2006 to 09/06/2006, with SMOSREX measurements, with a thirty minute frequency. It can be observed that the minimum of each curve is never found inside the interval.

In view of these results, a well-defined minimum within the prescribed boundaries of each parameter cannot be found. Reaching a minimum would require the parameters to visit unrealistic values. This can be caused by the model, the observations or our assimilation system. In order to explore the validity of the latter with SMOSREX forcing, a twin experiment was performed using the input forcing from SMOSREX. The conditions for the

twin experiment are the following: The ten most sensitive parameters to TB are optimized with their standard values. After the introduction of a random noise up to 50% their prior values, the assimilation of synthetic observations of TB is applied from 02/06/2006 to 09/06/2006, with a frequency of 30 minutes, using SMOSREX forcing.

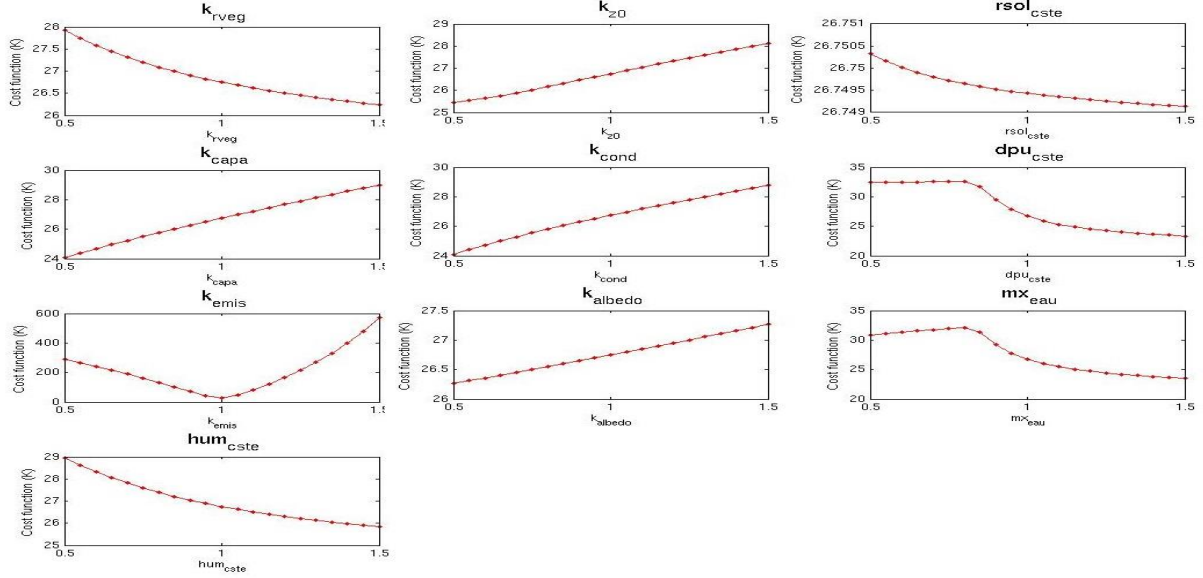


Figure 7. 17. Cost function variations around prior parameter values, from 02/06/2006 to 09/06/2006.

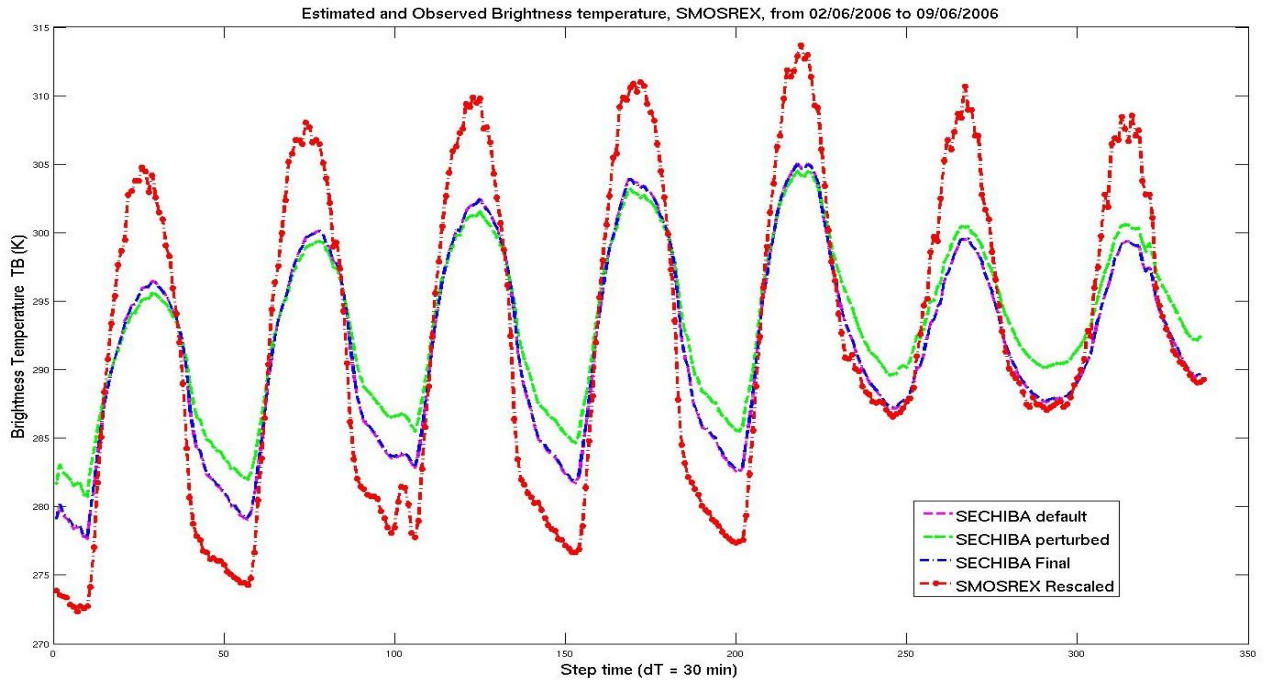
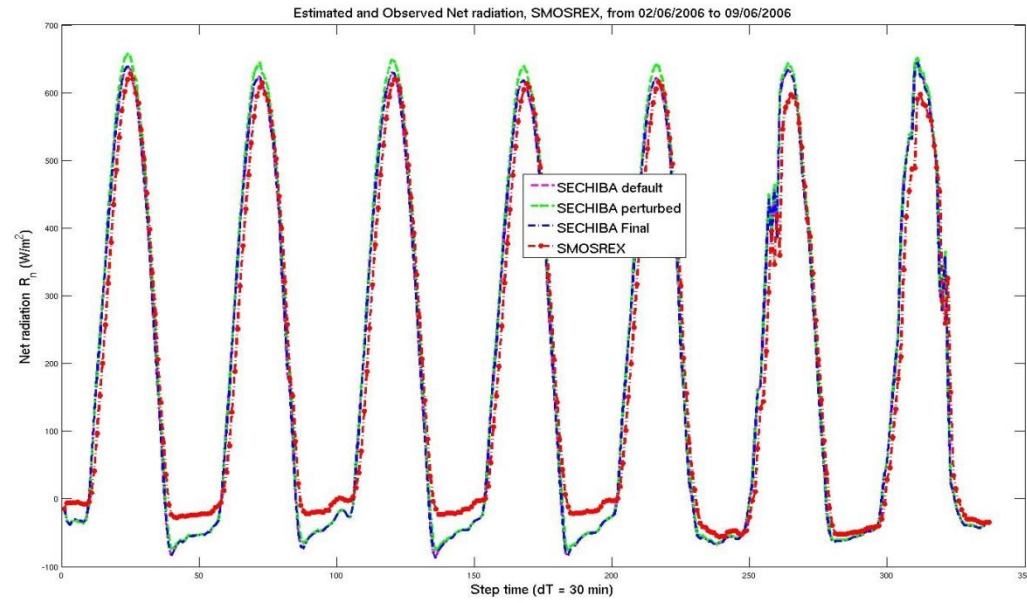
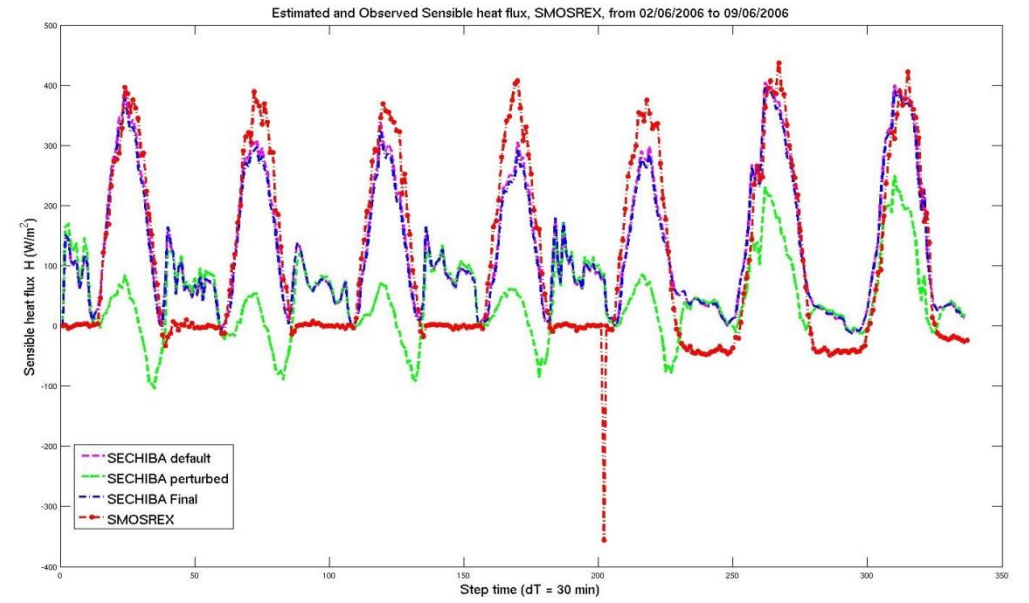


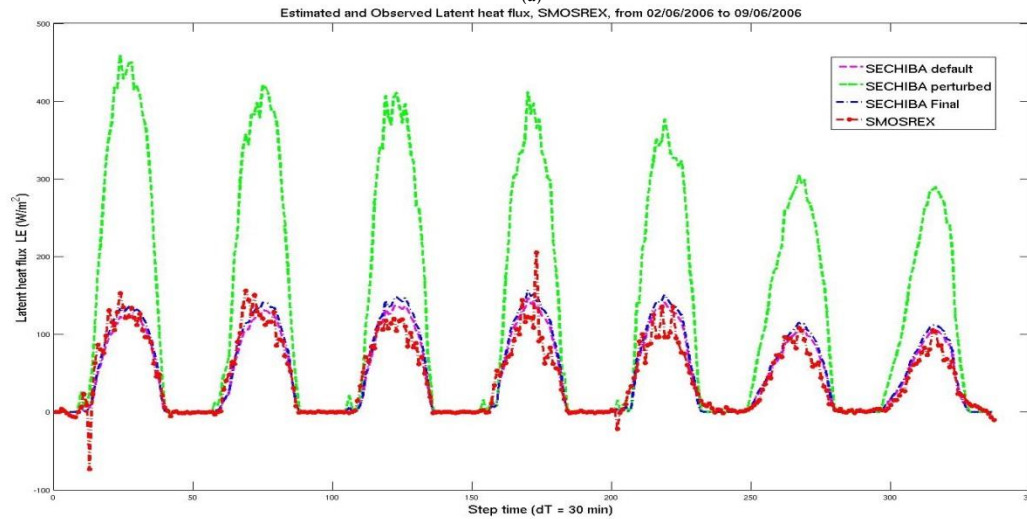
Figure 7. 18. Comparison of TB : observed (red), modeled with default parameter values (magenta), modeled with noised parameters (green) and posterior estimation using the latter as prior (blue), from 02/06/2006 to 09/06/2006.



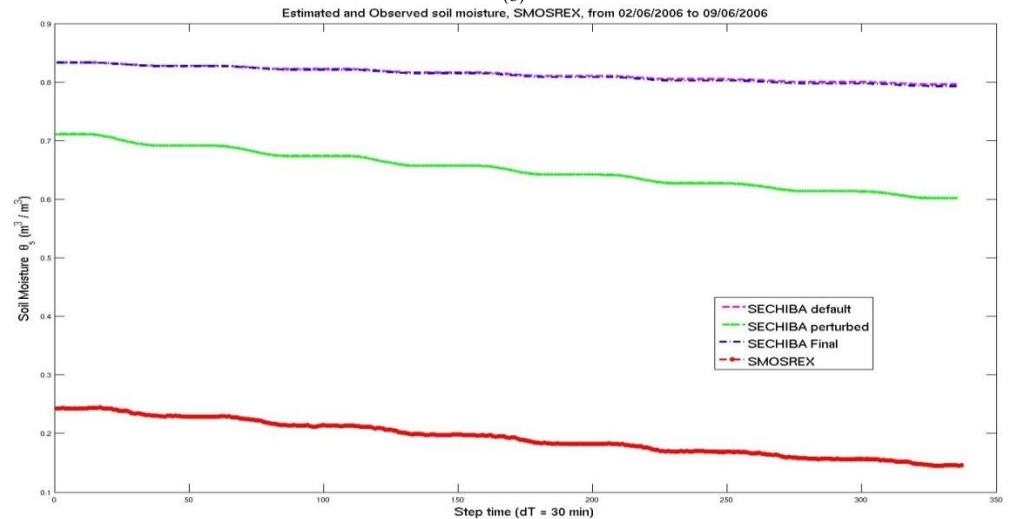
(a)



(b)



(c)



(d)

Figure 7. 19. Comparison of variables from SMOSREX observations (red), SECHIBA estimation with prior parameter values (magenta), SECHIBA perturbed (green) and SECHIBA estimations after twin experiment (blue), R_n (a), H (b), LE (c), θ_s (d).

We present the results of the twin experiments at the end of the assimilation in Fig.7.18 and Fig.7.19. The figures plot the different time series before and after perturbing the parameters; and the results of the twin experiment. Moreover, a comparison with the SMOSREX observations is provided. From Fig.7.18 and Fig.7.19, it can be observed that the parameter values that generated our standard estimations are retrieved, thus confirming the consistency of the assimilation system, as proven in Chapters 5 and 6. However, discrepancies in state variables and flux with rescaled SMOSREX observations remain large, with respect to the TB amplitude. The perturbations have contrasted impacts on the different variables. R_n is insensitive to the perturbation of the parameters (as observed in Fig.7.19.a). The curve before and after the perturbation are similar, although the perturbation can reach up to 50% of the initial value. So there is little hope that optimization of the selected set of parameter would improve this output. The LE flux is overestimated when parameters are perturbed. However the prior estimation of LE is closer to the SMOSREX observations. The soil moisture θ_s is overestimated with the parameter perturbation. The assimilation system retrieves the prior value.

These results allow us to state that the assimilation system is consistent in finding the optimal values in a twin experiment and confirms the parameter sensitivities and the usefulness of assimilating LST to constrain model parameters. The results confirm also that the bad performances obtained with real measurements are linked to the non-ability of our model to simulate a physical surface temperature comparable to measurements.

7.6 Conclusion

In this chapter, the assimilation of TB in SECHIBA was implemented. With different scenarios, including two different prior values of hum_{cste} (4 and 0.4) and two assimilation time windows (one day and seven days). Experiment results show that the model cannot reach a proper minimum to improve the variable estimation.

TB with CDF-Matching was presented, allowing us to match the statistics of SECHIBA outputs with the available measurements. Even though rescaling of the whole year 2006 helps to fit better our observations with CDF-Matching, the corrections are limited, since the observations may present time variability, biases and variance. If these discrepancies are limited to the directional effect of the measurements, this is not surprising, i.e. biases during summer or winter will not be similar. Instead, it may be better to apply CDF-Matching in a shorter time-period, involving the period assimilated with homogeneous climate conditions.

Comparing Experiments 1, 2 and 3 (where $hum_{cste} = 4$) with Experiments 4, 5 and 6 (where $hum_{cste} = 0.4$) it can be observed that the best results were obtained by controlling the six most sensitive parameters with an assimilation time window of a single day (Experiments 1 and 4). However, the choice of an arbitrary prior value is an initial approach. The next step would be to perform an exploration in the parameter, selecting the optimal prior values.

Even though variable estimation is slightly improved, this is due to final parameter values at the boundary of a variation interval in the cost function. A cost function sampling confirmed these results. We can observe that the cost function does not have a clear minimum in the ranges prescribed. In order to reach a minimum, it would require the parameters to visit unrealistic values. SECHIBA does not correctly simulate simultaneous temperature and fluxes and the relationship between the two is not always consistent according to the regime (or parameter values that are used). We must therefore work on the physics to improve the temperature simulation. Also, the parameter sensitivity to temperature is not always sufficient, giving as a result a flat cost function.

When attempting to calibrate the model with only a small number of parameters directly related to TB , the accuracy in estimating the state of the system can be either improved or degraded, depending on the prior scenario chosen for the experiment. However, improvements can be limited by several factors to consider: first, the parameters indirectly related to TB , which are involved in soil hydrology processes, have a role in the calculation of the energy budget; second, the model is not suitable to reproduce the physical reality of the measured phenomena. Indeed, observations proved impossible to fit in terms of realistic dynamics, showing the limitation of the assimilation method; finally, the control parameter vector size may be inadequate. Should other parameters be added to the initial list of eligible ones, thus allowing us to explore a wider spectrum of model responses?

Without prior information on the observations, the approach used in our assimilation scheme aims at reproducing the observations in the best way possible. This may not work very well due to model and observation errors. An improvement can be expected from introducing error information into the assimilation scheme, by using a weighted cost function. In the same way, it is necessary to evaluate the way thermodynamics and hydrology are coupled in this model, in order to introduce new physical processes, or perhaps to change the parameterization of the existing processes, especially the representation of LST .

Chapter 8

Conclusion and Perspectives

The objective of this thesis was to improve model variables of the ORCHIDEE biosphere model by constraining internal parameters with 4DVAR assimilation of land surface temperature observations. 4DVAR merges observations with the model, providing an optimal initial parameter estimate that improves the forecast potential of the model.

The first objective drove us first to the study of the model physics. The study of land surface models (as presented in Chapter 1) was a necessary introduction in order to understand the processes and variables that can be estimated with this type of models. In addition, specificities and complexities were identified. Afterwards, insights of ORCHIDEE and especially SECHIBA (which computes the energy and water budgets) were presented in Chapter 2. The equations related to the energy and water budgets, implemented in SECHIBA were included. Model components were identified in order to understand its dynamics, needed to build the modular graph in YAO. The FLUXNET and SMOSREX sites used in this work for providing *in situ* observations were presented. Both sites have measurements corresponding to SECHIBA forcing. Their measurements characteristics and different time series of input forcing and observed variables are shown, permitting the knowledge of the dynamics and particularities of each site.

Once the physical environment and data were presented, the components of our methodology of work were introduced: variational data assimilation concepts (Chapter 3) and the introduction to YAO (Chapter 4). The implementation of SECHIBA into YAO was explained in detail. It involved complex computational tasks: understanding SECHIBA original code, extracting important codes representing the model physics and finally using YAO meta-language to define the modular graph of the model and the internal procedures encapsulated in the different modules. Even if it was a laborious task, the implementation of

SECHIBA in YAO allows us to have an easily maintainable, modifiable and expandable code. It also offers the flexibility to run assimilation experiments with various scenarios in a very easy way. In addition, although in this work SECHIBA-YAO 1D was used in one dimension, the implementation was done to be used in two dimensions.

Once these steps were accomplished, we can say that SECHIBA-YAO 1D represents the original model accurately and that the adjoint model obtained with the backward integration of the modular graph, made with YAO, is an equivalent representation of the adjoint model.

Once SECHIBA-YAO 1D was validated, sensitivity analysis was performed (Chapter 5), obtaining a parameter hierarchy. Sensitivity analysis can indicate which parameters influence the most, the various output variables, and the accuracy with which it is necessary to estimate them. Sensitivities for FLUXNET sites were computed using land surface temperature as observation. The parameter hierarchy obtained with the sensitivity analysis serves as a basis to choose the most sensitive parameters of the considered variable. Furthermore, it gives an opportunity to better understand the dynamical behavior of the system, by highlighting over parametrization or parameters that do not concern observation operators.

The most sensitive parameters were then used in the different assimilation experiments performed in Chapter 6. Twin experiments were implemented using FLUXNET forcing, in order to test the robustness of our assimilation methodology. Several scenarios were tested with synthetic observations, noting the performances of the assimilation when changing the observation sampling, the assimilation time window, the observation type and the size of the control parameter set. For each experiment, parameter initial values were perturbed before the assimilation. The idea was to retrieve the initial value of the parameters that generates our synthetic observations. Results were promising, succeeding in most experiments to converge to the parameter prior values. Although results with twin experiments are limited, proving only that assimilation works accurately, they allow us to assess the impact of various configurations, in the parameter restitution performance.

The assimilation of *in situ* measurements was conducted in Chapter 7. The aim of this chapter was to assess the impact of assimilating land surface temperature observations. Different scenarios were tested: first by varying the size of the control parameter set, taking in some experiments the six or the ten most sensitive parameters. Second, by testing two different evapotranspiration regimes, prescribing different initial condition values for hum_{cste} (4 and 0.4) and finally with two assimilation time windows of one and seven days. All scenarios used *TB* measurements rescaled with CDF-Matching from SMOSREX database.

Sensitivity analysis was computed with this forcing since it is different from FLUXNET site, and the results obtained suggest that different processes are more or less predominant in each site.

When assimilating only TB , the scenarios with the best performance were with experiments controlling only the six most sensitive parameters (Experiments 1, 2, 4, and 5). In addition, using $hum_{cste} = 0.4$ improves flux estimation, as it can be seen in Experiments 5 and 6, compared to Experiments 2 and 3. It was observed that in all experiments, the model parameter final values are at the edge of the validity interval, thus indicating that the minimization process did not find a minimum within the defined intervals for each parameter. A cost function sampling confirmed these results. We can observe that the cost function does not have a clear minimum, in the ranges prescribed. Effectively reaching a minimum would require allowing the parameters to visit unrealistic values. SECHIBA does not correctly simulate simultaneously temperature and fluxes and the relationship between the two is not always consistent according to the regime (or parameter values that are used), giving as a result a flat cost function. We must therefore work on the physics to better simulate the temperature.

Results using two different prior values of hum_{cste} highlight the need to study the impact of this parameter in the model dynamics.

Several remarks stem from these results:

1. The assimilation system implemented to apply 4DVAR on SECHIBA is robust, since the resulting performances in the twin experiments are satisfactory for all sites (SMOSREX and FLUXNET).
2. It was noticed that the cost function, when assimilating real measurements from SMOSREX, does not have its minimum within the parameter intervals. It may be due to the non-ability of our model to simulate a physical surface temperature comparable to measurements.
3. In view of the results using a different prior value for the control parameter set, it was shown that the optimization of the surface temperature does not always induce improvements on surface fluxes. Depending on the parameter initial conditions, the physical processes and their consistency may vary, implying different results in the assimilation performances. Therefore, a preliminary work on model errors and sensitivity to initial conditions should be performed prior to the assimilation work, to correct model

biases and systematic errors.

4. The coupling between the hydrology and the thermodynamics in SECHIBA must be reviewed in order to reflect the interactions between these two parts of the model more accurately. In particular, a different hydrology in SECHIBA has been implemented, besides the Choisnel hydrology used in this work, reflecting more precisely the hydrologic processes, with eleven layers in the soil.
5. CDF Matching improves the observations. However it is important to apply it during a period where the climate conditions are homogeneous, which guarantees a better match of the variable dynamics between observations and estimations. In addition, the formula used to estimate TB can be reviewed to simulate it with higher accuracy.
6. An exhaustive study of the TB observations errors must be conducted also in order to retrieve more adapted weighing terms values that help us to assess accurately the measurement errors magnitude in the cost function, accounting in this way for the confidence we give to the observations. Finally, an extensive study of parameters correlations has to be done in order to obtain a wider comprehension of the model internal processes, accounting them in the cost function.

Perspectives

The work that should be undertaken to improve the performances obtained in this work is the following:

1. The different assimilation experiments performed with the SMOSREX dataset expose just preliminary results, showing only that local interactions may not be generalized to a regional study. A wider regional study, with different data sources, has to be done in order to account for the potential of variational data assimilation in constraining SECHIBA internal parameters.
2. The eleven layers hydrology procedure can be implemented in SECHIBA-YAO 1D, in order to constrain parameters by assimilating TB observations. Perhaps this hydrology would represent more accurately coupling to the soil thermodynamics.
3. Besides the parameters considered in the control set, there are other parameter of interest that can be considered in the assimilation process. Exploring the influence of the parameters in the estimations model variables is a mandatory task in order to estimate sensitivities and pertinence of parameters within the model. This exploration has to be

done in future works every time a new forcing is used or a different variable is observed (besides *TB*). With SECHIBA-YAO 1D, it is straightforward to perform this exploration: all parameters are already coded, the user just has to choose which parameters and in which variable the sensitivity are going to be computed.

4. Assimilate a different variable besides *TB*. For example soil moisture is a good candidate, since we observed that high discrepancies were obtained between estimations and SMOSREX measurements, thus incorporating relevant information that can be extracted from these observations, control parameters can be better constrained.

Challenges

After having implemented SECHIBA-YAO 1D, some reflections are made on implementation decisions taken and the way the work was performed. This is a constructive criticism on the work done. It should help making better decisions in the coming developments of models over YAO. When planning the implementation of a model using YAO, this scheme should be followed:

1. Definition and delimitation of the model: it defines the limits of the model to be implemented, which routines are vital and which can be eliminated from the original model. It must be made between the developer and the experts who know the model.
2. Adaptation with tools: the developer must first know YAO and the model to be implemented. The physics and dynamics must be understood to facilitate the validation process. If the model has already been implemented in a programming language, extra time is required to understand the logic that was used to produce the original code.
3. Modular graph: It is the responsibility of the developer to extract the dynamics of the model, its components and flow of inputs and outputs in order to precisely define the modular graph. This step is very important because it contains the essence of the model, which is its dynamics.
4. Direct model, adjoint model: the coding, based on the original model, is done module by module. At the end, a representation of the direct model and adjoint model is obtained.
5. Validation of the model: The direct model should be validated with the original model to confirm its accuracy. The adjoint model should be checked to verify its accuracy.

At early stages of the implementation, the approach based on the use of SECHIBA-Fortran routines was chosen, even though feasibility studies were not performed to see if the

code was suitable to a differentiation phase. Based on this decision, the general planning of the work was formulated, allotting little time to the coding phase, which turned to be the longest task. The reason why the original code could not be used is related to the logic used to produce this code: the inputs in the routines are systematically modified during the computing phase, making impossible to differentiate the inputs, the intermediate variables and the outputs. A neat code where input, internal processes and outputs are clearly defined and do not change during the calculations is necessary, in order to obtain a consistent code that can be differentiated. Since this was not the case for the version of SECHIBA-Fortran taken, the migration of the original code had to be done.

An extensive study of the code was made to identify external libraries used in the original model, which were not necessary in the implementation of SECHIBA-YAO. A wide comprehension of the use of these libraries was then done before the implementation of the code began, in order to implement possible process made by these libraries, used in the core of SECHIBA.

Once the coding phase has begun, the developer encountered a code that was undocumented (when the development began) making really hard to understand the dynamics of the different routines. Functions, temporal variables and constant did not have any explanation of their use. The developer systematically followed every input, output and process in order to understand the code dynamics.

Besides all these hurdles, understanding the logic used by the developer of SECHIBA was a crucial stage. The logic used to code in FORTRAN language is not always the same as the logic used in YAO meta-language. Just the fact of migrating a code is a laborious task, it is wide known that two different developers do not think in a similar way, reason why rewriting a code not only takes time but also may introduces errors.

References

- Aubinet, M., Vesala, T., Papale, D. Eddy Covariance: A Practical Guide to Measurement and Data Analysis. Springer Atmospheric Sciences Editions. 2012
- Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., Paw, K. T., Pilegaard, K., Schmid, H. P., Valentini, R., Verma, S., Vesala, T., Wilson, K., Wofsy, S. FLUXNET: a new tool to study the temporal and spatial variability of ecosystem-scale carbon dioxide, water vapor, and energy flux densities. Bull. Am. Meteorol. Soc., 82, pp. 2415–2434. 2001.
- Barrett, D., Renzullo, L. On the Efficacy of Combining Thermal and Microwave Satellite Data as Observational Constraints for Root-Zone Soil Moisture Estimation. CSIRO Land and Water, 1109-1127, Canberra, Australia. 2009.
- Bateni, S.M., Entekhabi, D., Jeng, D.S. Variational assimilation of land surface temperature and the estimation of surface energy balance components. Journal of Hydrology 481 (2013) 143–156. 2013.
- Brajard, J. Methodologie neuronale pour l'inversion des signaux satellitaires de couleur de l'océan. Traitements des aerosols absorbants et restitution de la concentration en chlorophylle-a. These de doctorat de l'Université Paris 6. 2006.
- Breierova, L., Choudhari, M. An Introduction to Sensitivity Analysis. Prepared for the MIT System Dynamics in Education Project. 1996.
- Brender, P. Modélisation des flux de carbone, d'énergie et d'eau entre l'atmosphère et les écosystèmes de steppe sahélienne, avec un modèle de végétation global. These de doctorat TELECOM ParisTech. 2012.
- Burba, G. Eddy Covariance Method, for Scientific, Industrial, Agricultural, and Regulatory Applications. LI-COR® Biosciences Editions. 2013.
- Castaing, W. Analyse de sensibilité et estimation de paramètres pour la modélisation hydrologique : potentiel et limitations des méthodes variationnelles. These de doctorat de l'Université Joseph Fourier - Grenoble I. 2007.
- Castelli F., Entekhabi, D., Caporali, E. Estimation of surface heat flux and an index of soil moisture using adjoint-state surface energy balance. Water Resources Research, vol. 35, no. 10, pages 3115-3125. 1999.
- Courault, D., Hadria, R., Ruget, F., Oliso, A., Duchemin, B., Hagolle, O. and Dedieu, G. Combined use of FORMOSAT-2 images with a crop model for biomass and water

- monitoring of permanent grassland in Mediterranean region. *Hydrol. Earth Syst. Sci.*, 14, 1731–1744. doi:10.5194/hess-14-1731-2010. 2010.
- De Rosnay P., Calvet J.-C., Kerr Y., Wigneron J.-P., Lematre F., Escorihuela M.J., Muoz-Sabater J., Saleh K., Barri J., Bouhours G., Coret L., Cherel G., Dedieu G., Durbe R., Fritz N.E.D., Froissard F., Hoedjes J., Kruszewski A., Lavenu F., Suquia D., Waldteufel P. : "SMOSREX: A Long Term Field Campaign Experiment for Soil Moisture and Land Surface Processes Remote Sensing" *Remote Sensing of Environment*, 102, 377-389, doi:10.1016/j.rse.2006.02.021. 2006.
- D'Orgeval, T. Impact du changement climatique sur le cycle de l'eau en Afrique de l'Ouest : Modelisation et incertitudes. These de doctorat de l'Université Paris 6. 2006.
- Ducoudré, N., Laval, K., and Perrier, A. SECHIBA, a new set of parametrizations of the hydrologic exchanges at the land/atmosphere interface within the LMD atmospheric general circulation model. *J. Climate*, 6(2) :248–273. 1993.
- François, C. , C. Ottlé and L. Prévot, Analytical parameterization of canopy directional emissivity and canopy directional radiance in the thermal infrared. Application on the retrieval of soil and foliage temperatures using two directional measurements. Part 1 : Theory, *Int. J. Remote Sensing*, 18,12, 2587-2621. 1997.
- Fréchet, M. Sur la loi de répartition de certaines grandeurs géographiques. *Journal de la Société de Statistiques de Paris* **82**: 114–122. 1941.
- Friedlingstein P, Joel G, Field C B, Fung I. Toward an allocation scheme for global terrestrial carbon models. *Global Change Biology*, 5, 755-770. 1999.
- Evensen, G. The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dyn.*, 53, 343–367. 2003.
- Gâteaux, R. Fonctions d'une infinité de variables indépendantes, *Bulletin de la Société Mathématique de France* 47: 70–96. 1919.
- Ghent, D., Kaduk, J., Remedios, J., Balzter, H. Data assimilation into land surface models: the implications for climate Feedbacks. *International Journal of Remote Sensing* Vol. 32, No. 3, 10 February 2011, 617–632. 2011.
- Giering, R., Kaminski., T. Recipes for Adjoint Code Construction. *ACM Transactions on Mathematical Software*, Vol. 24, No. 4, December 1998, Pages 437–474. 1998.
- Gilbert, J.C., Lemaréchal, C. Some numerical experiments with variable-storage quasi Newton algorithms, *Maths. Program*, 45, 407-435. 1989.
- Greenwald, T., Vukicevic, T., Grasso, L.,Voder Haar, T. Adjoint sensitivity analysis of an observational operator for visible and infrared cloudy-sky radiance assimilation. *Q. J. R. Meteorol. Soc.*, 130, pp. 685–705. 2004.
- Harrison, D.E., Chiodi A.M, Vecchi, G.A.Effects of surface forcing on the seasonal cycle of the eastern equatorial Pacific. *J. Mar. Res.* 67, 701-729. 2009.

- Hascoët, L., Pascual, V. The Tapenade Automatic Differentiation tool: principles, model, and specification. Research Report Nro 7557. Project Team Tropics. INRIA, France. 2012.
- Huang, C., Li, X., Lu, L. Retrieving soil temperature profile by assimilating MODIS LST products with ensemble Kalman filter. Cold and Arid Regions Environmental and Engineering Research Institute, CAS, Lanzhou, China. 2003.
- Järvinen, H. Observations and diagnostic tools for data assimilation. Meteorological Training Course Lecture Notes. European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, England. 1998.
- Kane, A. Assimilation De Données In Situ Et Satellitaires Dans Le Modèle De Biogéochimie Marine PISCES. These de doctorat de l'Université-de-Versailles-Saint-Quentin-en-Yvelines. 2010.
- Kirkby, M., Naden, P., Burt, T., Butcher, D. Computer simulation in physical geography, John Wiley & Sons, Chichester. 1992.
- Krinner, G., Viovy, N., Noblet-Ducoudre, N. de, Ogee, J., Polcher, J., Friedlingstein, P., Ciais, P., Sitch, S., and Prentice, I. C. A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochem. Cycles*, 19. 2005.
- Kuppel, S. Assimilation de mesures de flux turbulents d'eau et de carbone dans un modèle de la biosphère continentale. These de doctorat de l'Université-de-Versailles-Saint-Quentin-en-Yvelines. 2012.
- Kuppel, S. Assimilation de Mesures de Flux turbulents d'eau et de carbone dans un modèle de la Biosphère Continentale. These de doctorat de l'Université de Versailles Saint-Quentin-en-Yvelines, soutenue le 14 decembre 2012.
- Lahoz, W.; Khattatov, B. Data Assimilation Making Sense of Observations. Springer Editions. 2010
- Le Dimet, F.-X., Talagrand, O. Variational Algorithms for Analysis and Assimilation of Meteorological Observations: Theoretical Aspects. *J. Tellus, Series A, Dynamic Meteorology and Oceanography* 38. 1986.
- Liu, Y. and Gupta, H. V. Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework, *Water Resources.*, 43, W07401, doi:10.1029/2006WR005756. 2007.
- Louvel, S. Implementation of a Dual Variational Algorithm for Assimilation of Synthetic Altimeter Data in the Oceanic Primitive Equation Model micom. *J. Geophys. Res.*, 106(C5):9199–9212. 2001.
- Manrique-Sunen, A., Nordbo, A., Balsamo, G., Beljaars, A. Representing Land Surface Heterogeneity: Offline Analysis of the Tiling Method. *Journal Of Hydrometeorology*, 14, 850-868. 2013.
- Musy, A., Higy, C. Hydrologie 1: Une science de la nature. Presses Polytechniques et Universitaires Romandes, Lausanne, Suisse. 2003.

- Musy, A., Soutter, M. Physique du Sol. Collection Gérer L'environnement. Presses Polytechniques et Universitaires Romandes, Lausanne, Suisse. 1991.
- Nardi, L. Formalisation et automatisation de YAO, générateur de code pour l'assimilation variationnelle de données. These de doctorat du Conservatoire National d'Arts et Metiers. 2011.
- Nardi, L., Sorrow, C., Badran, F., and Thiria, S. *YAO: A Software for Variational Data Assimilation Using Numerical Models*. Computational Science and its Applications - ICCSA 2009. International Conference, 5593, 2, 621-636. 2009.
- NOAA National Weather Service Jetstream. Available on: http://www.education.noaa.gov/Freshwater/Water_Cycle.html.
- Olioso, A., Inoue, Y., Ortega-Farias, S., Demarty, J., Wigneron, J.P., Braud, I., Jacob, F., Lecharpentier, P., Otter, C., Calvet, J. C., and Brisson, N.: Future directions for advanced evapotranspiration modeling: assimilation of remote sensing data into crop simulation models and SVAT models, *Irrig. Drainage Syst.*, 19, 377–412, 2005.
- Parrens, M. Assimilation des données SMOS dans un modèle des surfaces continentales : mise en oeuvre et évaluation sur la France. These de doctorat de l'Université de Toulouse, 2013.
- Parrens, M., Mahfouf, J., Barbu, A., Calvet, J. Assimilation of surface soil moisture into a multilayer soil model: design and evaluation at local scale. Manuscript prepared for *Hydrol. Earth Syst. Sci. Discuss.* 2013.
- Pipunic, R. C., Walker, J. P., and Western, A. Assimilation of remotely sensed data for improved latent and sensible heat flux prediction: A comparative synthetic study. 19th International Congress on Modelling and Simulation, Perth, Australia. 2008.
- Pires, C., Vautard, R. and Talagrand, O. On extending the limits of variational assimilation in nonlinear chaotic systems. *Tellus* 48A, 96–121. 1996.
- Polcher, J., McAvaney, B., Viterbo, P., Gaertner, M.-A., Hahmann, A., Mahfouf, J.-F., Noilhan, J., Phillips, T., Pitman, A.J., Schlosser, C.A., Schulz, J.-P., Timbal, B., Verseghy D., and Xue, Y. A proposal for a general interface between land-surface schemes and general circulation models. *Global and Planetary Change*, 19 : 263-278. 1998.
- Reichle, R. H. and Koster, R. D.: Bias reduction in short records of satellite soil moisture, *Geophys. Res. Lett.*, 31, doi:10.1029/2004GL020938, <http://dx.doi.org/10.1029/2004GL020938>, 2004.
- Reichle, R., Kumar, S., Mahanama, S., Koster, R., Liu, Q.. Assimilation of satellite-derived skin temperature observations into land surface models. *Journal of Hydrometeorology*, 11, 1103-1122. 2010.
- Ridler, M., Sandholt, I., Butts, M., Lerer, S., Mougin, E., Timouk, F., Kergoat, L., Madsen, H. Calibrating a soil–vegetation–atmosphere transfer model with remote sensing

- estimates of surface temperature and soil surface moisture in a semi arid environment. *Journal of Hydrology* 436–437 (2012) 1–12. 2011.
- Robert, C, Blayo, E., Verron, J. Comparison of reduced-order, sequential and variational data assimilation methods in the tropical Pacific Ocean. *Ocean Dynamics* 56, 5-6 (2006) 624-633. 2007.
- Rump, S. How reliable are results of computers, *Jahrbuch Uberblicke Mathematik*, pp. 163-168. 1983.
- Saltelli, A. *Sensitivity Analysis*. John Wiley & Sons Edition. USA. 2008
- Scipal, K., Holmes, T., de Jeu, R., Naeimi, V., and Wagner, W.: A possible solution for the problem of estimating the error structure of global soil moisture data sets, *Geophys. Res. Lett.*, 35, doi:10.1029/2008GL035599, <http://dx.doi.org/10.1029/2008GL035599>, 2008.
- Sitch, S., Smith, B., Prentice, I.C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J.O., Levis, S., Lucht, W., Sykes, M.T., Thonicke, K., Venevsky, S. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. *Glob. Change Biol.* 9, 161 –185. 2003.
- Svendsen, H. "The effect of clear sky radiation on crop surface temperature determined by thermal thermometry." *Agricultural and Forest Meteorology* 50: 239-243, 1990.
- Thiria, S., Badran, F., Sorrow, C. Yao: Un logiciel pour les modeles numeriques et l'assimilation de donnees. Technical report, LOCEAN-IPSL. 2006.
- Wainwright, J., Mulligan, M. *Environmental Modeling: Finding Simplicity in Complexity*. John Wiley & Sons Edition, USA. 2004.

Figure Index

FIGURE 1. 1 WATER MOVES THROUGH THE EARTH, CHANGING STATE AND DRIFTING TO THE ATMOSPHERE, THE OCEANS AND OVER THE LAND SURFACE AND UNDERGROUND, IN DIFFERENT PROCESSES THAT COEXIST. IT IS SUBJECT TO COMPLEX PROCESSES; AMONG THEM WE CITE PRECIPITATION, EVAPORATION, TRANSPIRATION, INTERCEPTION, RUNOFF, INFILTRATION, PERCOLATION, STORAGE AND SUBSURFACE FLOWS. THESE VARIOUS MECHANISMS ARE MADE POSSIBLE BY THE INCOMING SURFACE ENERGY. (SOURCE: NOAA NATIONAL WEATHER SERVICE JETSTREAM).	15
FIGURE 1. 2 SOIL PROFILE AND THE DIFFERENT TYPES OF SATURATION ZONES.	17
FIGURE 1. 3 THERMAL CONDUCTION BETWEEN TWO SOLID (LEFT) AND THERMAL CONVECTION BETWEEN A SOLID AND A FLUID (RIGHT). (SOURCE: MUSY AND SOUTTER, 1991)	18
FIGURE 1. 4 GLOBAL EARTH ENERGY BUDGET. (SOURCE: NASA)	21
FIGURE 1. 5 DAYLIGHT AND NIGHT ENERGY BUDGETS. (SOURCE: MUSY ET SOUTTER (1991)).	23
FIGURE 2. 1 ENERGY BALANCE	35
FIGURE 2. 2 SPECIFIC VARIABLES INVOLVED IN HYDROLOGICAL BUDGET COMPUTING, AS IN D'ORGEVAL (2006).	37
FIGURE 2. 3 REPRESENTATION OF HORIZONTAL AIR FLOW THAT PASSES THROUGH THE TOWER, AND CONSISTS OF DIFFERENTLY SIZED EDDIES. (SOURCE BURBA, 2013).	41
FIGURE 2. 4 THE AIR FLOW CONSIST OF ROTATING EDDIES. EDDY 1 MOVES PARCEL OF AIR C_1 DOWN WITH THE SPEED W_1 , WHILE EDDY 2 MOVES PARCEL C_2 UP WITH THE SPEED W_2 . (SOURCE BURBA, 2013)	41
FIGURE 2. 5 DISTRIBUTION OF TOWER SITES (SOURCE HTTP://FLUXNET.ORNL.GOV)	42
FIGURE 2. 6 TIME SERIES OF FORCING VARIABLES FOR FLUXNET HARVARD FOREST FROM 1995 TO 1996.	43
FIGURE 2. 7. TIME SERIES OF FORCING VARIABLES FOR FLUXNET KRUGER PARK FROM 2002 TO 2003.	44
FIGURE 2. 8 TIME SERIES OF FORCING VARIABLES (A) AND MEASURED FLUXES (B) FOR SMOSREX, FROM 2005 TO 2006.	46
FIGURE 3. 1 BASIC ITERATION OF VARIATIONAL DATA ASSIMILATION.	50
FIGURE 3. 2 PART OF A MODULAR GRAPH WITH 3 MODULES. ONLY ONE BASIC CONNECTION IS ALLOWED BY A GIVEN INPUT, BUT MULTIPLE BASIC CONNECTIONS CAN EMERGE FROM THE SAME OUTPUT. SOURCE: NARDI ET AL, 2009	52
FIGURE 3. 3 LEVELS IN A MODULAR GRAPH. F1 AND F2 ARE INPUT MODULES, F3 AND F4 ARE INTERNAL MODULES AND F5 IS AN OUTPUT MODULE. SOURCE: NARDI ET AL, 2009.	53
FIGURE 3. 4 TWO GRAPH ABSTRACTION LEVELS. AT THE LOWER LEVEL (A), WE BUILD THE GRAPH Γ_g ; AT THE SPACE LEVEL (B), THE SAME GRAPH Γ_g IS REPEATED FOR EACH GRID POINT (2D IN THIS EXAMPLE). THE SPACE CONNECTIONS BETWEEN THE Γ_g GRAPHS CORRESPOND TO THE BASIC CONNECTIONS BETWEEN THE MODULES. SOURCE: NARDI ET AL, 2009	55

FIGURE 4. 1 YAO INPUT AND OUTPUT COMPONENTS, FROM NARDI (2011). USER SPECIFIES THE <i>MODULES</i> , <i>DESCRIPTION</i> AND <i>CHAPEAU</i> FILES. YAO COMPILER GENERATES THE <i>EXECUTABLE</i> AND THROUGH THE <i>INSTRUCTION</i> FILE, ASSIMILATIONS CAN BE LAUNCHED UNDER SPECIFIC SCENARIOS	61
FIGURE 4. 2 INPUT AND OUTPUT DATA FLOW THROUGH YAO. USERS CAN UPLOAD OBSERVATIONS AND MODEL STATES TO THE PROJECT ENVIRONMENT, FOLLOWING SPECIFIC YAO NOTATION.	62
FIGURE 4. 3 SECHIBA SUBROUTINES AND ITS CORRESPONDING OUTPUTS.....	64
FIGURE 4. 4 SECHIBA HYPER GRAPH, SHOWING GENERAL MODEL DYNAMICS	65
FIGURE 4.5 COMPARISON BETWEEN SECHIBA-FORTRAN AND SECHIBA-YAO 1D OUTPUTS, USING FLUXNET HARVARD FOREST FORCING FROM 19/07/2003 TO 23/07/2003. EACH CURVE IS A TIME SERIES DURING 4 DAYS FOR 8 DIAGNOSTIC VARIABLES. RED CURVES ARE THE ESTIMATIONS COMPUTED WITH SECHIBA-FORTRAN. BLUE CURVES REPRESENTS SECHIBA-YAO -1D ESTIMATION. (A) IS WITH BARE SOIL AND (B) WITH GRASSLAND.	70
FIGURE 4.6 COMPARISON BETWEEN SECHIBA-FORTRAN AND SECHIBA-YAO 1D OUTPUTS, USING FLUXNET KRUGER PARK FORCING FROM 07/02/2003 TO 11/02/2003. EACH CURVE IS A TIME SERIES DURING 4 DAYS OF 8 DIAGNOSTIC VARIABLES. RED CURVES ARE THE ESTIMATIONS COMPUTED WITH SECHIBA-FORTRAN. BLUE CURVES REPRESENTS SECHIBA-YAO -1D ESTIMATION. (A) IS WITH BARE SOIL AND (B) WITH GRASSLAND.	71
FIGURE 4. 7 RESULTS OF THE ADJOINT TEST APPLIED TO SECHIBA-YAO 1D. THE DIFFERENCE BETWEEN BOTH TERMS IS PRESENTED IN THE LOWER PART OF THE FIGURE.	73
FIGURE 4. 8 GRADIENT TEST RESULT USING KRUGER PARK FORCING FROM 07/02/2003 TO 11/02/2003, WITH AND WITHOUT CADNA WITH GRASSLAND (PFT 11). RESIDUE CURVE (BLUE) MUST VARY AS A^2 (BLACK DOTTED LINE) SO THAT THE TEST IS VALID.	75
FIGURE 7. 1 SUB-GRAPH ADDED TO THE MODULAR GRAPH OF SECHIBA-YAO 1D IN ORDER TO SIMULATE TB AND COMPARE IT, TO SMOSREX TB OBSERVATIONS.	114
FIGURE 7. 2 COMPARISON BETWEEN SIMULATED (BLUE) AND OBSERVED (GREEN) TB , FROM 02/06/2006 TO 09/06/2006. ABSCISSA CORRESPONDS TO TIME STEPS (336 FOR A WEEK) AND ORDINATE CORRESPONDS TO BRIGHTNESS TEMPERATURE IN KELVIN.....	115
FIGURE 7. 3 SCATTER PLOT OF SIMULATED TB VS. OBSERVED TB , FOR 2006. ABSCISSA CORRESPONDS TO ESTIMATED TB AND ORDINATE CORRESPONDS TO SMOSREX MEASUREMENTS OF TB , DURING 2006.	115
FIGURE 7. 4 COMPARISON OF THE CUMULATIVE DISTRIBUTION FUNCTION (CDF) OF THE BRIGHTNESS TEMPERATURE (TB) ESTIMATED WITH SECHIBA (BLUE), OBSERVED IN SMOSREX (GREEN) AND RESCALED SMOSREX AFTER CDF-MATCHING (RED), FOR THE YEAR 2006.	116
FIGURE 7. 5 TIME SERIES FROM 02/06/2006 TO 09/06/2006 FOR TB ESTIMATED BY SECHIBA (BLUE), OBSERVED IN SMOSREX (GREEN) AND RESCALED AFTER CDF-MATCHING FOR SMOSREX MEASUREMENTS (RED).	117
FIGURE 7. 6 COMPARISON BETWEEN SIMULATED (θ_s) AND OBSERVED (θ_s^{obs}) SOIL MOISTURE, FROM 02/06/2006 TO 09/06/2006, USING SMOSREX MEASUREMENTS.....	118

FIGURE 7. 7 GRADIENT SENSITIVITY ANALYSIS USING SMOSREX FORCING, WITH PFT 11, THE 02/06/2003. THE SENSITIVITIES WERE COMPUTED ON TB . CURVES REPRESENT TB DERIVATIVE WITH RESPECT TO EACH PARAMETER GIVEN BY THE ADJOINT MODEL, EVERY HALF HOUR OVER A DAY.....	119
FIGURE 7. 8. PRIOR AND POSTERIOR ASSIMILATION RMSE VALUES FOR MODEL VARIABLES. ABSCISSA CORRESPONDS TO THE DAY OF THE ASSIMILATION, BEGINNING FROM THE 02/06/2006 TO 02/07/2006. ORDINATE CORRESPONDS TO RMSE VALUES. RED CURVES STAND FOR THE RMSE OF THE DIFFERENT VARIABLES PRIOR TO ASSIMILATION. BLUE CURVES ARE THE RMSE AFTER ASSIMILATION.	121
FIGURE 7. 9 PARAMETERS POSTERIOR VALUES, WHEN ASSIMILATING BRIGHTNESS TEMPERATURE. PLOTS ARE THE CUMULATIVE HISTOGRAMS OF THE PARAMETER POSTERIOR VALUES OBTAINED AFTER ONE DAY ASSIMILATION.	122
FIGURE 7. 10. TIME SERIES FOR TB BEFORE AND AFTER ASSIMILATION. RESULTS CORRESPOND TO EXPERIMENT 2. RED CURVES ARE THE RESCALED SMOSREX MEASUREMENTS. GREEN CURVES ARE SECHIBA ESTIMATIONS WITH THE PARAMETERS PRIOR VALUES. BLUE CURVES ARE THE ESTIMATED TB WITH THE OPTIMIZED PARAMETERS.	123
FIGURE 7. 11. TIME SERIES FOR R_N (A) H (B), LE (C) AND θ_s (D) BEFORE AND AFTER ASSIMILATION, FOR EXPERIMENT 2, FROM 02/06/2006 TO 09/06/2006. RED CURVES ARE THE SMOSREX MEASUREMENTS. GREEN CURVES ARE ESTIMATIONS WITH THE PARAMETERS PRIOR VALUES. BLUE CURVES ARE THE VARIABLES WITH THE OPTIMIZED PARAMETERS.	125
FIGURE 7. 12. COMPARISON BETWEEN SECHIBA TB WITH $HUM_{CSTE}=0.4$ (MAGENTA), SECHIBA TB WITH $HUM_{CSTE}=4$ (BLUE) AND SMOSREX ORIGINAL MEASUREMENTS OF TB (GREEN), FROM 02/06/2006 TO 09/06/2006, USING SMOSREX FORCING. ABSCISSA CORRESPONDS TO TIME STEPS (336 FOR A WEEK) AND ORDINATE CORRESPONDS TO BRIGHTNESS TEMPERATURE IN KELVIN.	127
FIGURE 7. 13. CUMULATIVE DISTRIBUTION FUNCTION (CDF) OF THE BRIGHTNESS TEMPERATURE (TB) ESTIMATED WITH SECHIBA (BLUE), SMOSREX (GREEN) AND RESCALED SMOSREX (RED).	127
FIGURE 7. 14. COMPARISON BETWEEN SIMULATED (θ_s) AND OBSERVED SOIL MOISTURE (θ_s^{obs}), FROM 02/06/2006 TO 09/06/2006, USING SMOSREX MEASUREMENTS.	128
FIGURE 7. 15. GRADIENT SENSITIVITY ANALYSIS USING SMOSREX FORCING, WITH PFT 11, THE 02/06/2003. THE SENSITIVITIES WERE COMPUTED ON THE TB , USING $HUM_{CSTE}=0.4$. CURVES REPRESENT TB DERIVATIVES WITH RESPECT TO EACH PARAMETER GIVEN BY THE ADJOINT MODEL.	129
FIGURE 7. 16. (A) COMPARISON OF RMSE VALUES PRIOR AND AFTER ASSIMILATION, DURING A MONTH. RED CURVES STAND FOR THE RMSE OF VARIABLES PRIOR TO ASSIMILATION. BLUE CURVES ARE THE RMSE AFTER ASSIMILATION. (B) PARAMETERS POSTERIOR VALUES, WHEN ASSIMILATING TB . PLOTS ARE THE CUMULATIVE HISTOGRAMS OF THE PARAMETER POSTERIOR VALUES OBTAINED AFTER ONE DAY ASSIMILATION.	131
FIGURE 7. 17. COST FUNCTION VARIATIONS AROUND PRIOR PARAMETER VALUES, FOR 02/06/2006 TO 09/06/2006 TIME-WINDOW.	134
FIGURE 7. 18. COMPARISON OF TB : OBSERVED (RED), MODELED WITH DEFAULT PARAMETER VALUES (MAGENTA), MODELED WITH NOISED PARAMETERS (GREEN) AND POSTERIOR ESTIMATION USING THE LATTER AS PRIOR (BLUE).....	134

FIGURE 7. 19. COMPARISON OF VARIABLES FROM SMOSREX OBSERVATIONS (RED), SECHIBA ESTIMATION WITH PRIOR PARAMETER VALUES (MAGENTA), SECHIBA PERTURBED (GREEN) AND SECHIBA ESTIMATIONS AFTER TWIN EXPERIMENT (BLUE), R_N (A), H (B), LE (C), θ_s (D). 135

Table Index

TABLE 2. 1 ORCHIDEE'S PLANT FUNCTIONAL TYPE DESCRIPTION (D'ORGEVAL 2006).....	33
TABLE 2. 2 ORCHIDEE'S PLANT FUNCTIONAL TYPE PRINCIPAL PARAMETERS (D'ORGEVAL 2006).....	34
TABLE 2. 3 INPUT VARIABLES RECEIVED BY SECHIBA	35
TABLE 2. 4 ENERGY BUDGET FLUXES.....	37
TABLE 2. 5HYDROLOGICAL BUDGET FLUXES	38
TABLE 2. 6 SECHIBA PARAMETERS STUDIED IN THIS WORK. THERE ARE 6 INTERNAL PARAMETERS, INVOLVED IN THE MODEL ESTIMATIONS AND 5 MULTIPLYING FACTORS THAT ARE IMPOSED TO SPECIFIC FLUXES	39
TABLE 4. 1SECHIBA SUBROUTINES AND NUMBER OF MODULES	67
TABLE 4. 2 RMSE AND RELATIVE ERROR BETWEEN SECHIBA-FORTRAN AND SECHIBA-YAO 1D. THESE ERRORS WERE COMPUTED BASED ON A YEAR SIMULATION ON BOTH MODELS.....	69
TABLE 4. 3 COST FUNCTION TEST RESULTS GIVEN BY YAO.....	73
TABLE 6. 1 SCENARIO PROPERTIES.....	102
TABLE 6. 2 SCENARIOS FOR EACH OF THE 6 EXPERIMENTS.	103
TABLE 6. 3 SAMPLING FREQUENCIES FOR EXPERIMENT 1	104
TABLE 6. 4 VARIABLE RMSE AND CONTROL PARAMETER RELATIVE ERRORS PRIOR AND POSTERIOR TO THE ASSIMILATION, FOR EXPERIMENT 1 (DIFFERENT FREQUENCIES SAMPLING IN THE OBSERVATIONS), USING KRUGER PARK FORCING, PFT 11, FROM 11/02/2003 TO 11/03/2003.	104
TABLE 6. 5 VARIABLE RMSE AND CONTROL PARAMETER RELATIVE ERRORS PRIOR AND POSTERIOR TO ASSIMILATION, FOR EXPERIMENT 1 (DIFFERENT FREQUENCIES SAMPLING IN THE OBSERVATIONS), USING HARVARD FOREST FORCING, PFT 11, FROM 28/08/1996 TO 28/09/1996.	105
TABLE 6. 6 VARIABLES RMSE AND CONTROL PARAMETER RELATIVE ERRORS PRIOR AND POSTERIOR TO ASSIMILATION, FOR EXPERIMENT 2 (DIFFERENT AMPLITUDE OF RANDOM NOISE IN THE OBSERVATION), USING KRUGER PARK FORCING, FROM 11/02/2003 TO 18/02/2003.....	106
TABLE 6. 7 VARIABLES RMSE AND CONTROL PARAMETER RELATIVE ERRORS PRIOR AND POSTERIOR TO ASSIMILATION, FOR EXPERIMENT 2 (DIFFERENT AMPLITUDE OF RANDOM NOISE IN THE OBSERVATION), USING HARVARD FOREST FORCING, FROM 28/08/1996 TO 04/09/1996.....	106
TABLE 6. 8SURFACE VARIABLES RMSE AND PARAMETER RELATIVE ERRORS BEFORE AND POSTERIOR TO THE ASSIMILATION PROCESS FOR EXPERIMENT 3 TO 5 (DIFFERENT CONTROL PARAMETER SET), USING KRUGER PARK FORCING, FROM 11/02/2003 TO 18/02/2003.	107
TABLE 6. 9 SURFACE VARIABLE RMSE AND PARAMETER RELATIVE ERRORS BEFORE AND POSTERIOR TO THE ASSIMILATION PROCESS FOR EXPERIMENTS 3 TO 5 (DIFFERENT CONTROL PARAMETER SET), USING HARVARD FOREST FORCING, FROM 28/08/1996 TO 28/08/1996.....	108
TABLE 7. 1 SECHIBA PARAMETERS STUDIED IN THIS WORK. THERE ARE FIVE INTERNAL PARAMETERS, INVOLVED IN THE MODEL ESTIMATIONS AND SIX MULTIPLYING FACTORS THAT ARE IMPOSED. IN ADDITION THE PRIOR RANGE IS SPECIFIED FOR EACH PARAMETER, DEFINING THE PRESCRIBED INTERVAL OF VARIATIONS.....	113
TABLE 7. 2. PARAMETER HIERARCHY USING SMOSREX FORCING, FOR BRIGHTNESS TEMPERATURE.	120

TABLE 7.3 CONFIGURATION FOR EXPERIMENT 1. THE SIX MOST SENSITIVE PARAMETERS IN THE CALCULATION OF TB ARE OPTIMIZED. THE PARAMETER VALUES ARE SET TO THEIR PRIOR VALUES (TABLE 7.1).....	120
TABLE 7. 4. SCENARIOS FOR EXPERIMENTS 2 AND 3. BOTH EXPERIMENT USED RESCALED TB OBSERVATIONS, WITH A 30 MINUTES SAMPLING.....	122
TABLE 7. 5. PRIOR AND POSTERIOR RMSE FOR EXPERIMENTS 2 AND 3. GRAY CELLS IN THE TABLE SHOW THE BEST RMSE VALUES FOR EACH SCENARIO.	123
TABLE 7. 6. PARAMETER POSTERIOR VALUES FOR EXPERIMENTS 2 AND 3, AFTER ASSIMILATING RESCALED TB . PRIOR VALUES FOR EACH PARAMETER ARE NORMALIZED TO 1. GRAY CELLS SHOW PARAMETER AT THE BOUNDARIES.	124
TABLE 7. 7. PARAMETER HIERARCHY USING SMOSREX FORCING, FOR BRIGHTNESS TEMPERATURE	130
TABLE 7. 8. EXPERIMENT 4. THE SIX MOST SENSITIVE PARAMETERS OF TB ARE OPTIMIZED. THE PARAMETER VALUES ARE SET TO THEIR PRIOR VALUES, EXCEPTING HUM_{CSTE} , WHICH IS EQUAL TO 0.4.	130
TABLE 7. 9. SCENARIOS FOR EXPERIMENTS 5 AND 6. EVERY EXPERIMENT USED RESCALED TB OBSERVATIONS, WITH A 30 MINUTES SAMPLING. HUM_{CSTE} IS EQUAL TO 0.4.	132
TABLE 7. 10. ASSIMILATION PERFORMANCES WHEN ASSIMILATING RESCALED USING $HUM_{CSTE} = 0.4$. BY COMPARING BOTH EXPERIMENTS, GREY CELLS IN THE TABLE SHOW THE BEST RMSE VALUE FOR EACH SCENARIO.	132
TABLE 7. 11. PARAMETER POSTERIOR VALUES FOR EXPERIMENTS 5 AND 6, AFTER ASSIMILATING RESCALED TB WITH $HUM_{CSTE} = 0.4$. PRIOR VALUES FOR EACH PARAMETER ARE NORMALIZED TO 1. GRAY CELLS SHOW PARAMETER VALUES AT THE LIMIT OF THEIR PRIOR RANGE.	133

Appendix A

Representing the tangent linear and adjoint of a modular through a modular graph

A.1. Tangent Linear of a modular graph

Hereafter, for ease of presentation and understanding, we use the matrix notation, the calculation results can be obtained without explicitly express the Jacobean matrices (i.e. using automatic generators). It is assumed that in a modular graph Γ from a model M , we can compute the tangent linear of each module F_q , based on a perturbation \mathbf{dx}_q (in a \mathbf{x}_q point). There by, we note \mathbf{F}_q the matrix corresponding to the tangent linear equal to the Jacobean matrix of F_q calculated in \mathbf{x}_q , the corresponding module output perturbation is equal to $\mathbf{dy}_q = \mathbf{F}_q \mathbf{dx}_q$. In the same way that the input vector \mathbf{x}_q , the vector \mathbf{dx}_q comes from perturbations in the outputs \mathbf{dy}_p . The tangent linear model can be computed by a forward propagation in the modular graph similar to the forward procedure. An example of this calculation is given in Figure A.1.

(2) Lin_forward algorithm

This algorithm is applied after all inputs \mathbf{x}_q , from all modules have been calculated (for this, we apply once the forward algorithm for an input vector \mathbf{x} defined in the external context of the modular graph).

- 1) Initialize perturbation \mathbf{dx} from input vector \mathbf{x} (this corresponds to all basic connections receiving data from the external environment).
- 2) Browse the module in the sense of the topological order. Calculate, in a way similar to the procedure forward, their linear tangent $\mathbf{dy}_q = \mathbf{F}_q \mathbf{dx}_q$ at the point \mathbf{x}_q .
- 3) Retrieve the result as a vector \mathbf{dy} formed by perturbations transmitted to the external environment

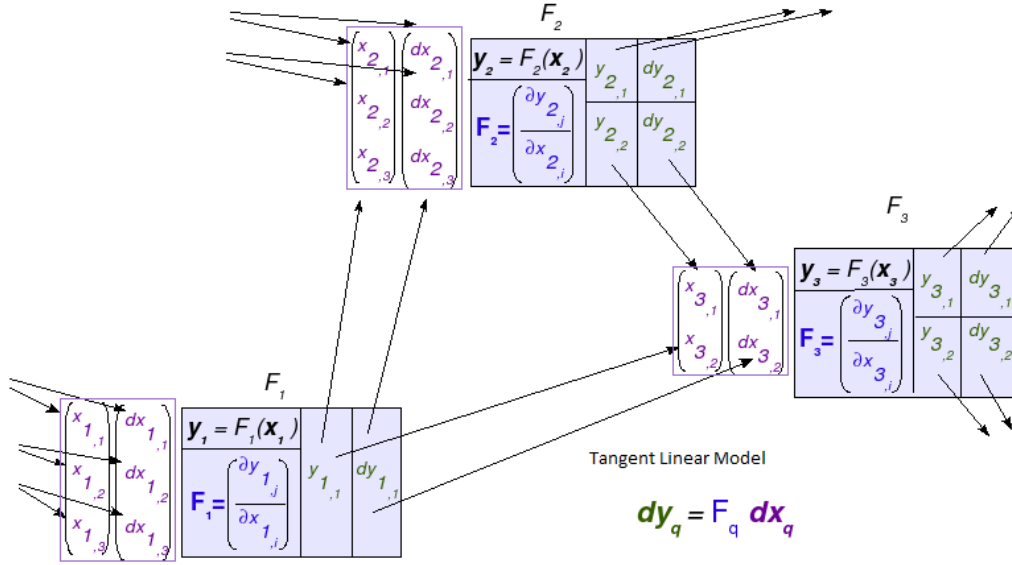


Figure A.1. Calculation of tangent linear in a modular graph. F_q is a module for which the vector \mathbf{x}_q (with components $x_{q,i}$) has already been calculated, and a perturbation \mathbf{dx}_q (with components $dx_{q,i}$). We compute $\mathbf{dy}_q = \mathbf{F}_q \mathbf{dx}_q$ which corresponds to the calculation of the tangent linear of the module F_q . $dy_{q,j}$ components can in turn be used by successor modules to form their own components $dx_{q,i}$ on the graph. Source: Nardi et al, 2009

A.2. Adjoint of a modular graph

As for the tangent linear model, we assume that we can calculate the adjoint model of each module. Thus, for a module F_p having an input vector \mathbf{x}_p , when receiving a vector \mathbf{dy}_p with the same size as the output vector, its adjoint model in \mathbf{x}_p is represented by $\mathbf{dx}_p = \mathbf{F}_p^T \mathbf{dy}_p$, having the same dimension as the input vector of F_p . The matrix \mathbf{F}_p^T is the transpose of the Jacobean matrix of the module F_p calculated at the point \mathbf{x}_p . We then demonstrate (see Appendix 1) that we can calculate the adjoint model of the global model M (represented by its graph) for each vector \mathbf{dy} having the same size as the output vector of the model ($\mathbf{dx} = \mathbf{\Gamma}^T \mathbf{dy}$). To do so, we compute the graph in reverse topological order, corresponding to evaluate the arcs in the opposite direction, this process it's called back propagation of the modular graph (backward).

The path in backward mode presents some difficulties compared to the path in the forward mode. Indeed, in forward mode, data transmission is always done through the basic connections. As illustrated in Fig.A.2.a, one and only one data is transmitted to a particular entry. Contrary, when passing in backward mode, more data can be transmitted to a particular output of module F_p (example y_{p2} the Fig.A.2.a). In the case of the adjoint, an intermediate

calculation is required. It consists in preparing the vector \mathbf{dy}_p of module F_p before calculating $\mathbf{dx}_p = \mathbf{F}_p^T \mathbf{dy}_p$. It is proved that the j^{th} component of \mathbf{dy}_p must be calculated as the sum of all data types of $dx_{q,i}$, transmitted to it by back propagation along all basic connections taking their input into y_{pj} . Thus, in the case of Fig.A.2.a, we must calculate $dy_{p2} = dx_{q2} + dx_{11}$. Different $dx_{p,i}$ back propagated are either derived from a calculation by a successor module F_p or they are calculated (or initialized) by the external context of the modular graph.

Finally, adjoint calculation of \mathbf{dx} in the global model ($\mathbf{dx} = \mathbf{M}^T \mathbf{dy}$) is form by vector components of \mathbf{dx}_p . These operations are summarized in Algorithm 3. Fig.A.4 shows an example of the calculations involved.

(3) Backward algorithm	
This algorithm is applied after all inputs \mathbf{x}_p of all modules have been calculated one time with the forward algorithm for an input vector \mathbf{x} .	
1) Initialize the output vector \mathbf{dy} of the modular graph Γ (corresponding to all basic connections transmitting data to the exterior environment).	
2) Visit the modules in reverse topological order (backward of the graph). For each module F_p , calculate \mathbf{dy}_p (as described in this paragraph) and calculate the adjoint $\mathbf{dx}_p = \mathbf{F}_p^T \mathbf{dy}_p$. This adjoint is computed in \mathbf{x}_p (we assume by hypothesis that this calculation is possible).	
3) Retrieve the adjoint of \mathbf{dx} result. This vector is constituted by the components of the vectors \mathbf{dx}_p corresponding to basic connections taking their inputs in the exterior environment	

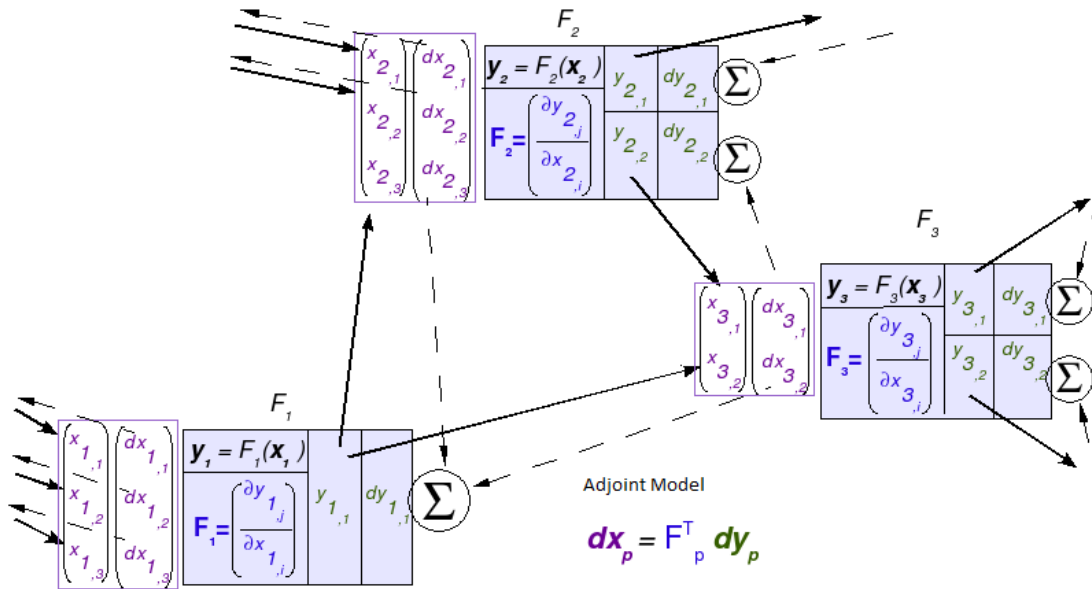


Figure A.2. Adjoint calculation in a modular graph. The dotted arrows represent the path in backward mode during which the adjoint calculation is performed. For this calculation, however, we recall that each component dy_{pj} from a module consists of the successors $dx_{q,i}$

sum of these modules. This operation is represented by the Σ symbol on the graph. Source: Nardi et al, 2009

In Fig.A.2 the technique used to calculate the adjoint method in a modular graph is presented. This is the same graph as that of Fig.A.4. The notations are equivalent except that the index p is replaced by q .

Both algorithms (2 and 3), assume that we know how to calculate the tangent linear and the adjoint for each module F_p . Modules can have very different complexities. In a simple case, where the module is an analytic function, we can explicitly calculate the Jacobean matrix of F_p and make the product $F_p \mathbf{dx}_p$ and $F_p^T \mathbf{dy}_p$. On more complex modules, we can use programs that do these calculations (i.e. a program obtained after using an automatic generator, even by another modular graph).

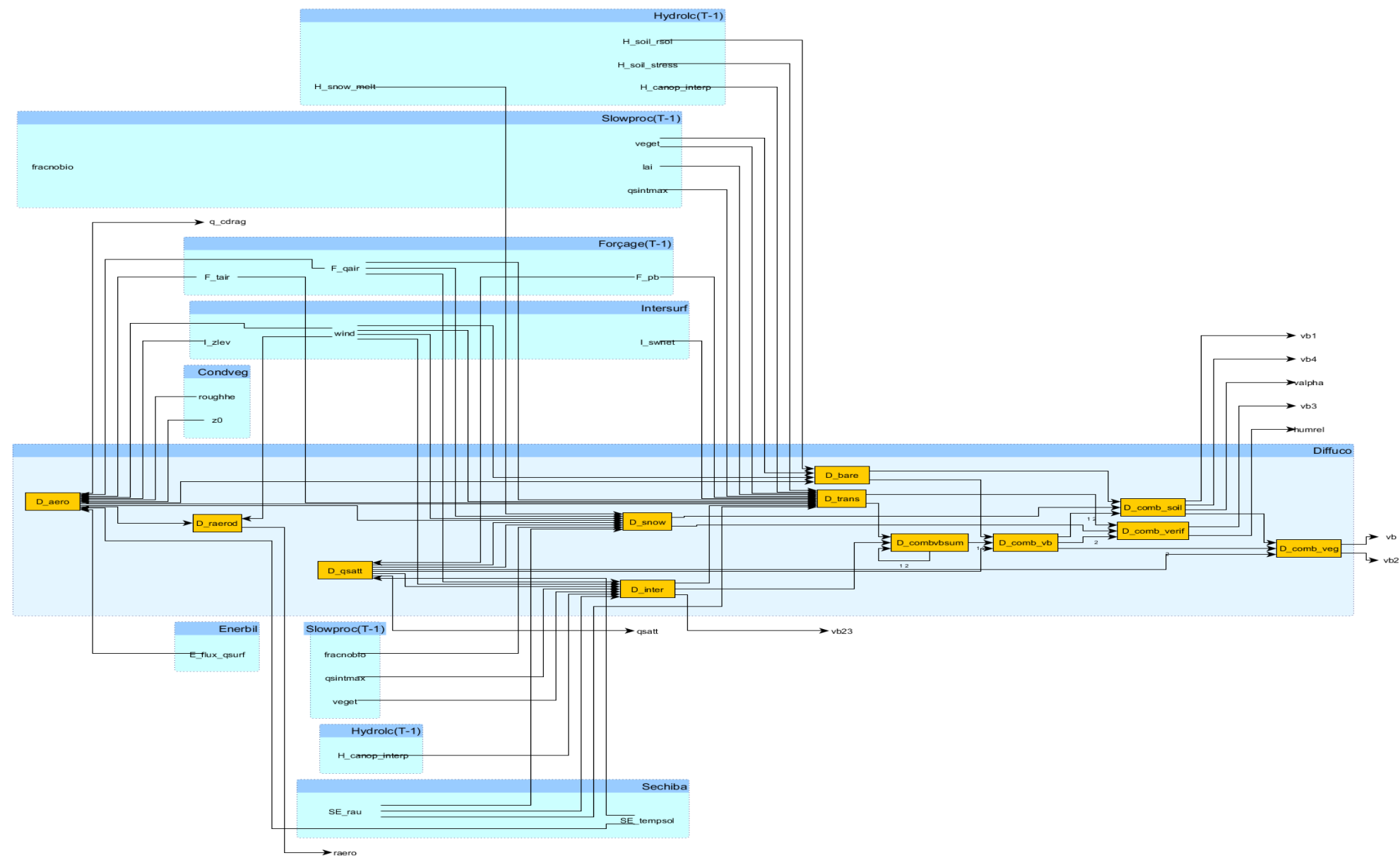
Appendix B

Modular graphs for each major routine of SECHIBA

B.1. Outline

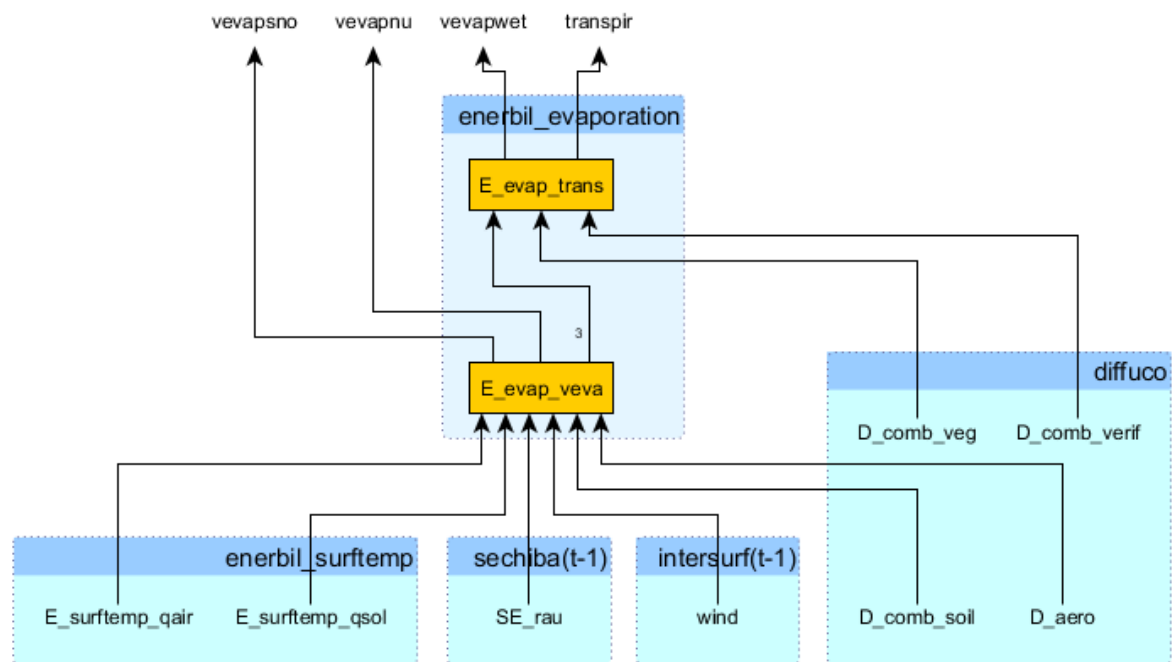
In this section, the modular graphs corresponding to each of the most important routines of SECHIBA are presented. As a reminder these routines are: DIFFUCO (diffusion coefficients), ENERBIL (energy budget), HYDROLIC (hydrologic budget), SLOWPROC (slow processes such as LAI update), THERMOSOIL (soil thermodynamics), CONDVEG (vegetation conditions). In addition, general internal calculations are grouped in INTERSURF (general parameters calculation).

B.2. DIFFUCO

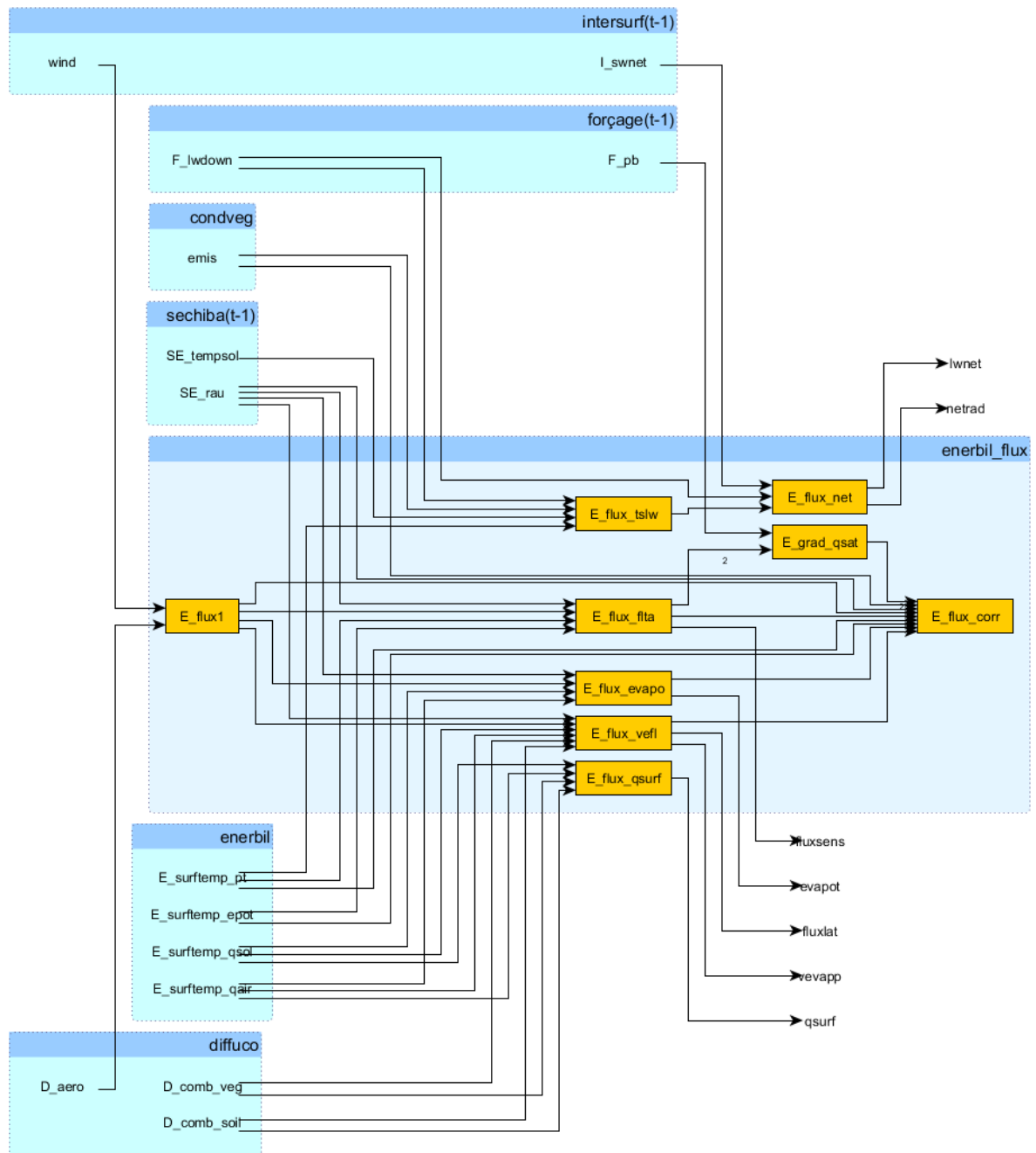


B.3. ENERBIL for evaporation

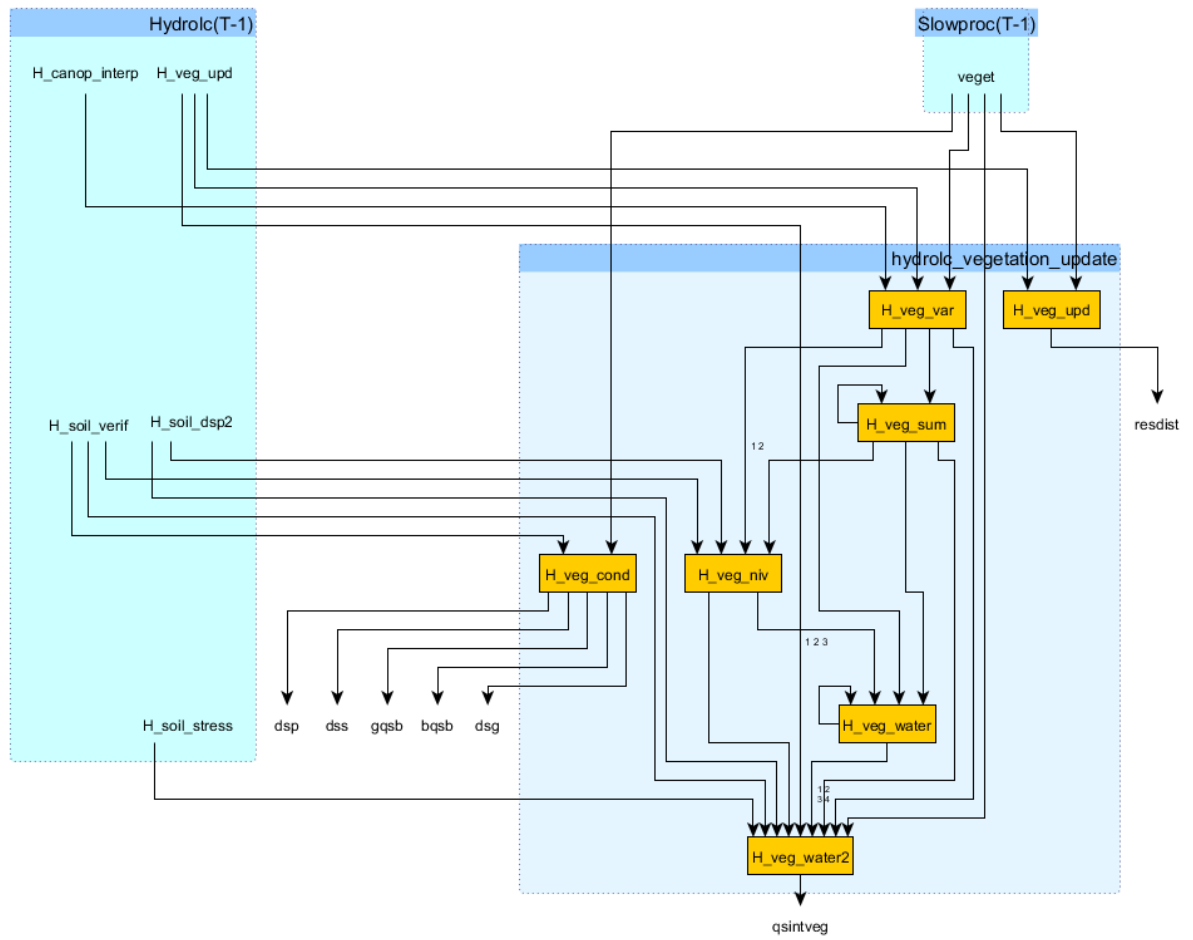
This graph represents the evaporation processes modeled inside ENERBIL



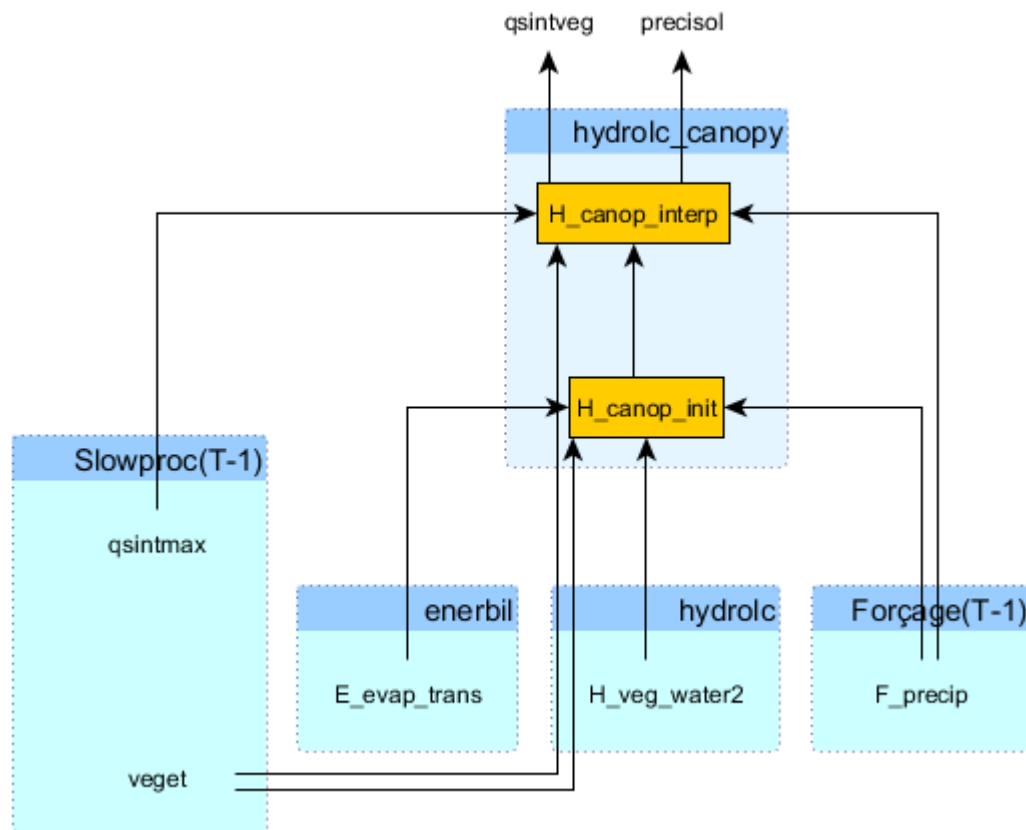
B.4. ENERBIL for surface fluxes



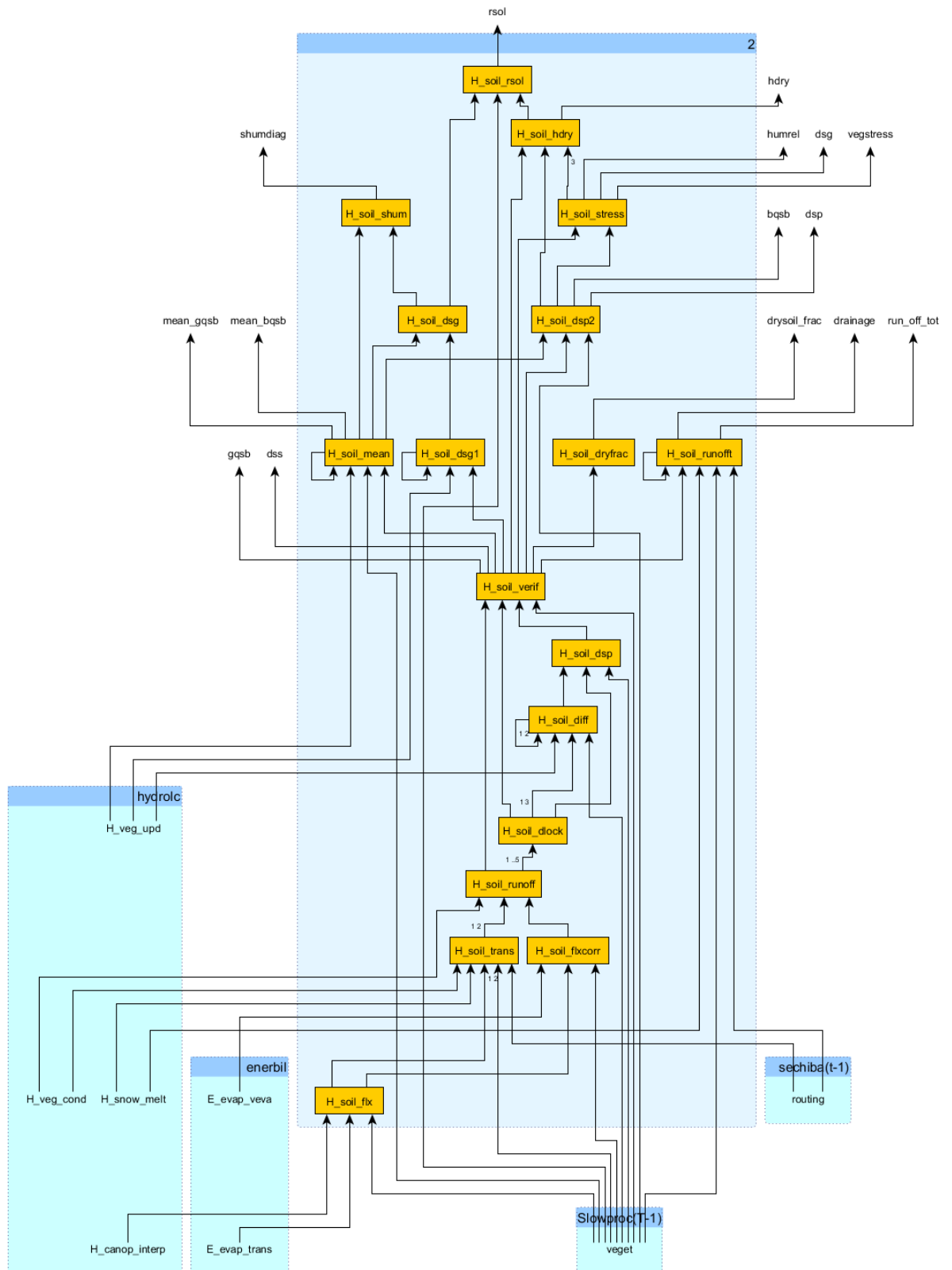
B.5. HYDROLc for soil water content



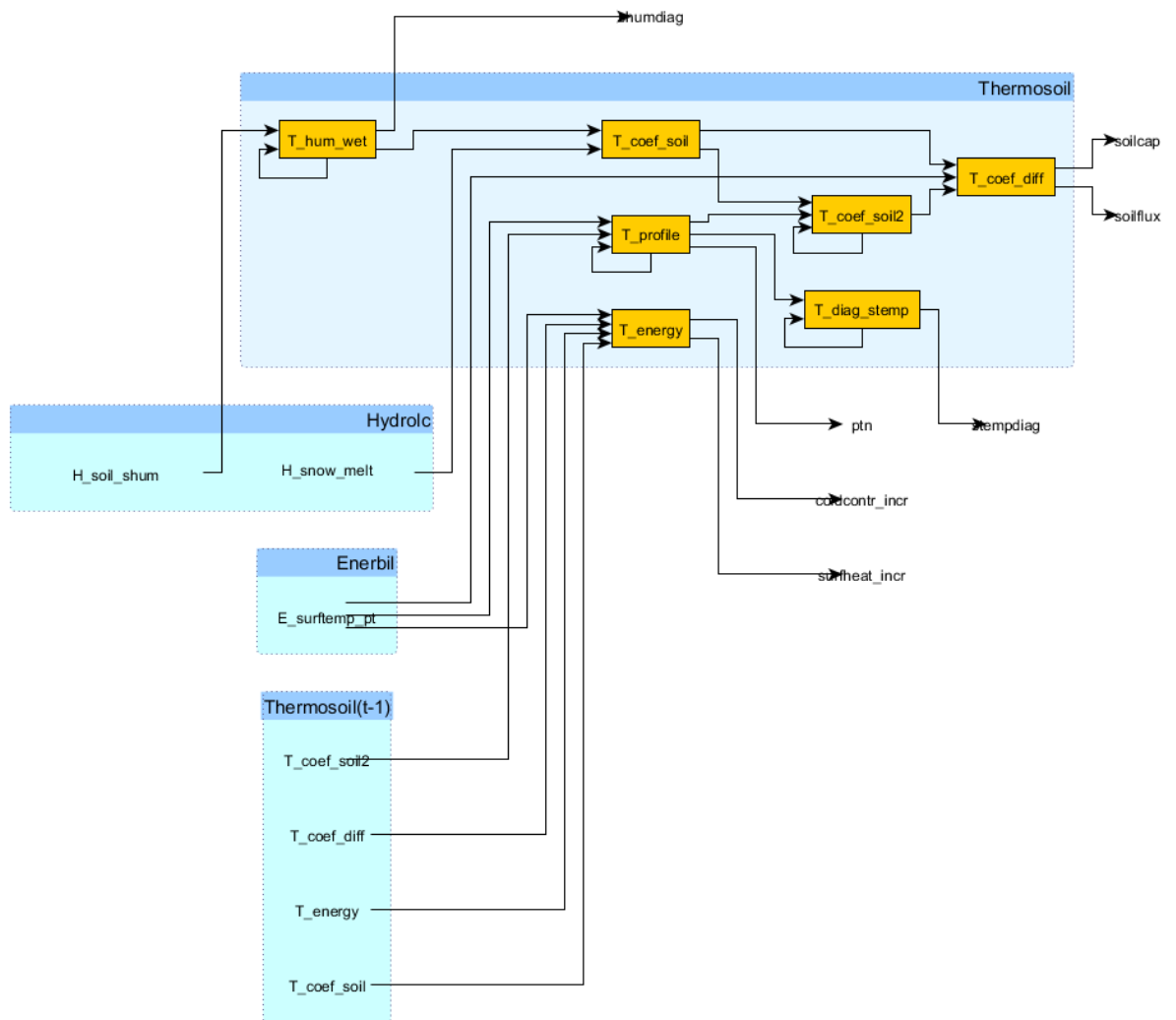
B.6. HYDROLc for canopy water interception



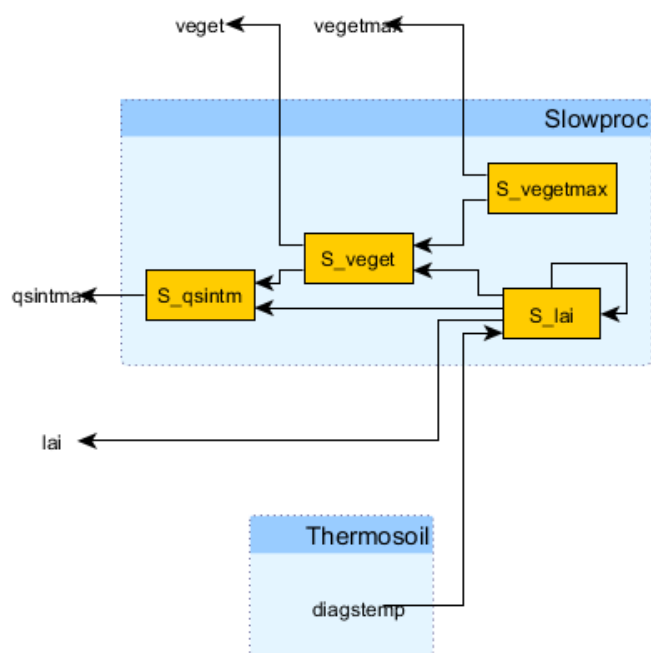
B.7. HYDROLIC for soil moisture update



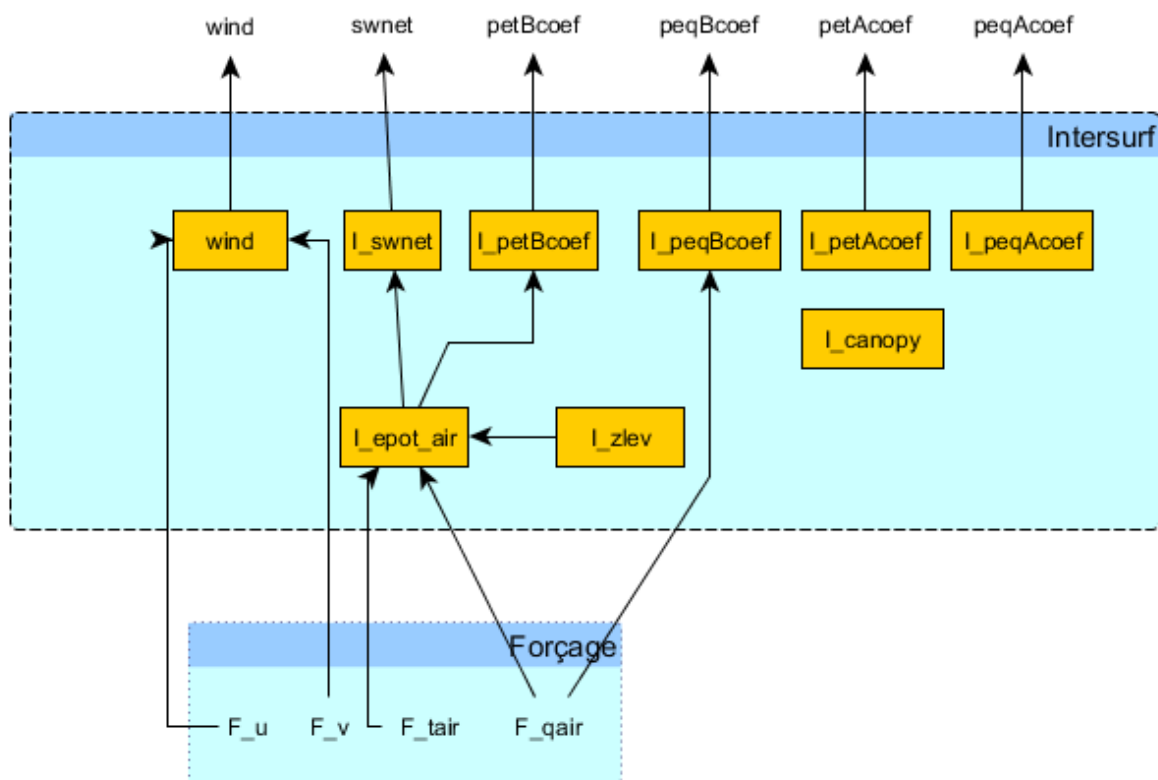
B.8. THERMOSOIL for soil temperature update



B.9. SLOWPROC



B.10. INTERSURF



Appendix C

Sensitivity Analysis using a Cost Function

C.1 Introduction

Complementary to the sensitivity analysis, another approach to perform a sensitivity analysis in the model parameters consists in the definition of a cost function, measuring the misfit between the simulated and observed temperature. This approach only requires the use of the direct model (like the finite differences). The cost function is defined by

$$J = \sum_{i=0}^t (M(x)_i - obs_i)^2 \quad (5.1)$$

where $M(x)$ is the simulated model output and obs_i is the observed measurement. We generate first, in a period of one week, synthetic observations of obs with the prior value of the parameter. Next, a sampling of the cost function is made, by performing consecutives calculation of $M(x)$ with a variation interval for each parameters, defined as $\pm 50\%$ the true parameter value.

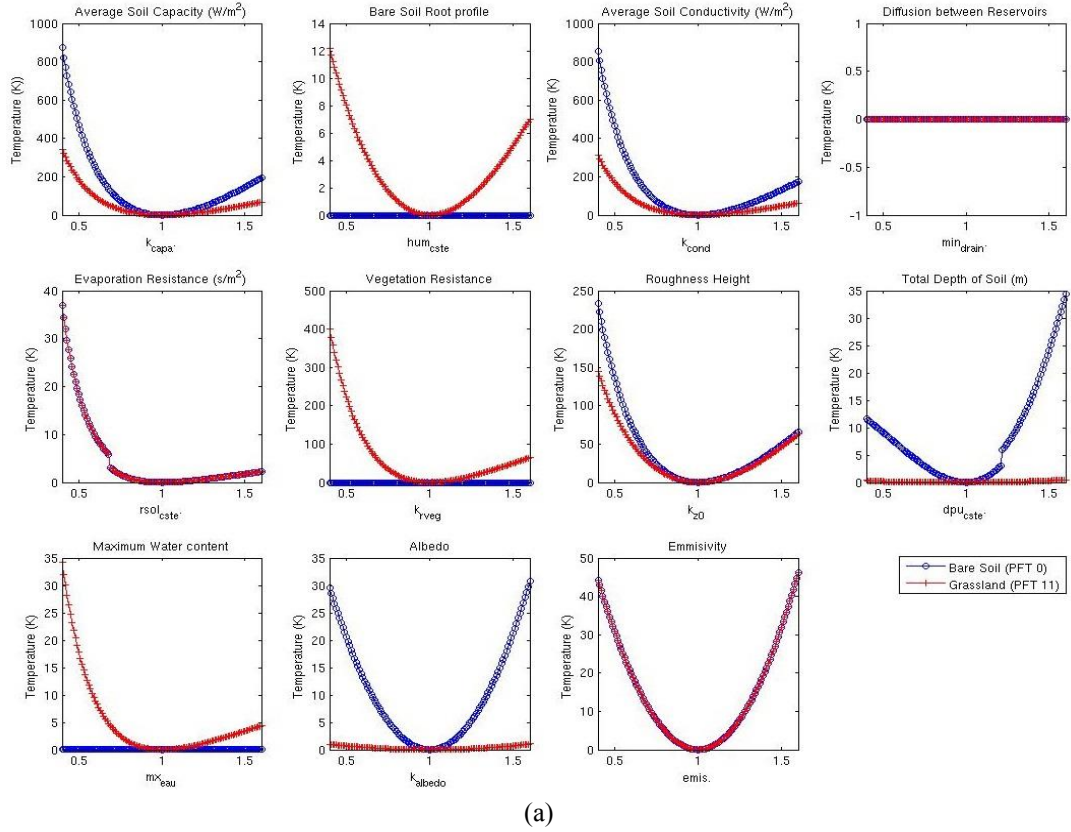
The approach using a cost function is very useful to perform a sensitivity analysis, if an adjoint model is not available. It is straightforward: given the direct model and the target state variables a cost function computing can be easily done to rapidly have an idea of the parameter influence in the model output. The result using a cost function can be used as a priori information to estimates sensitivities with the adjoint model.

C.2. Land surface Temperature sensitivity

This sensitivity analysis is made during a week, for soil temperature using both FLUXNET sites (Kruger Park and Harvard Forest) with PFT 0 and 11 and SMOSREX (PFT 11).

The curves presented in Figure C.1 represent the cumulated square difference of

simulated and observed temperatures, for both sites, with bare soil and grassland. Given the perturbation made for each parameter, the sampling gives us an idea of the influence of the variation of a single parameter has on the soil temperature estimation. When introducing the vegetation layer, several parameters have a higher impact in the soil temperature estimation. Only k_{albedo} and $rsol_{cste}$ curves remain constant.



(a)

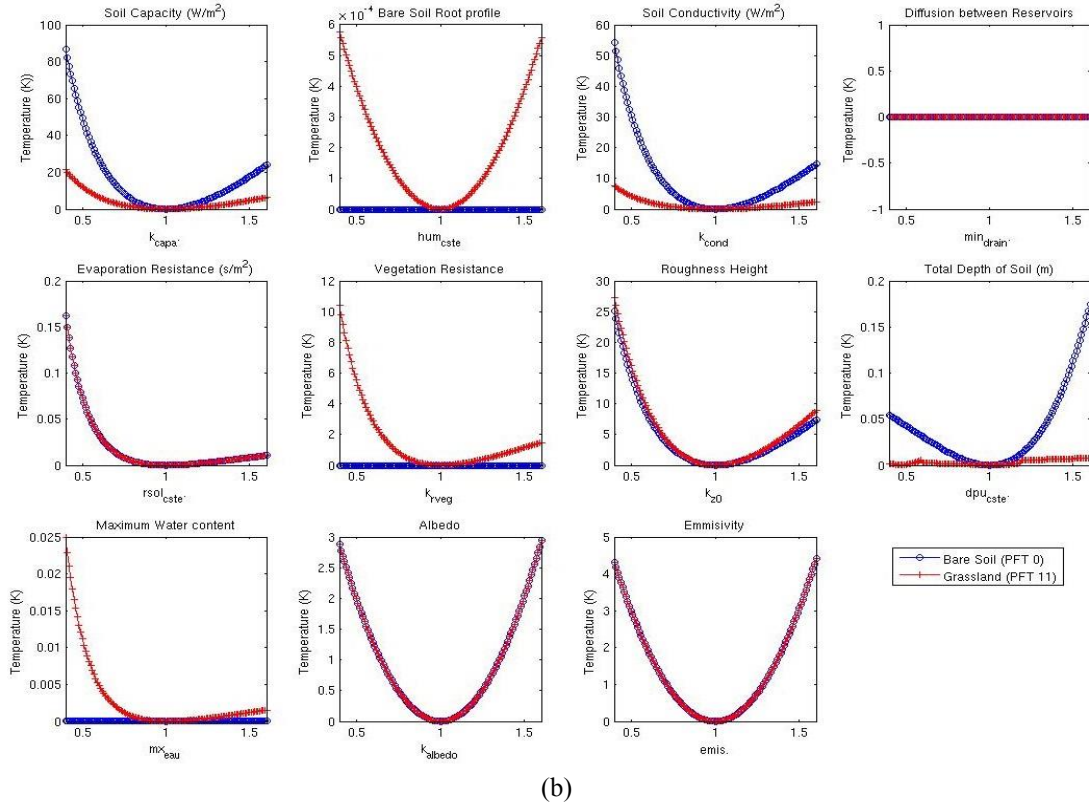


Figure C.1. Cost Function sensitivity analysis during a month for soil temperature with PFT 0 (blue curves) and PFT 11 (red curves) in FLUXNET Harvard Forest (a), from 24-08-1996; and Kruger Park (b), from 11-02-2003.

For the others, in the presence of grassland, the physics of the model introduced new variables affecting the sensitivity of these parameters. The 4 most influential parameters, for bare soil, are: k_{z0} , $rsol_{cste}$, k_{cond} and k_{capa} ; and for grassland, the hierarchy is: $kz0$, $krveg$, K_{cond} and k_{capa} . Both sites have the same hierarchy, showing that these parameters are independent from the forcing and can have values generalizing various types of vegetation, soils and climates. In addition, this hierarchy corresponds to the sensitivity analysis results performed in the previous section. However, in the Kruger Park results, it can be seen that the magnitude of the cost function is higher than the Harvard Forest results, as expected from the previous results. This is due to the forcing and the specific characteristics of each site. Harvard Forest is more stable during the days of testing. The shortwave solar radiation and the wind speed are smoother in this site than the Kruger Park.

Similar to the study made with FLUXNET sites, the cost function analysis was performed using SMOSREX forcing.

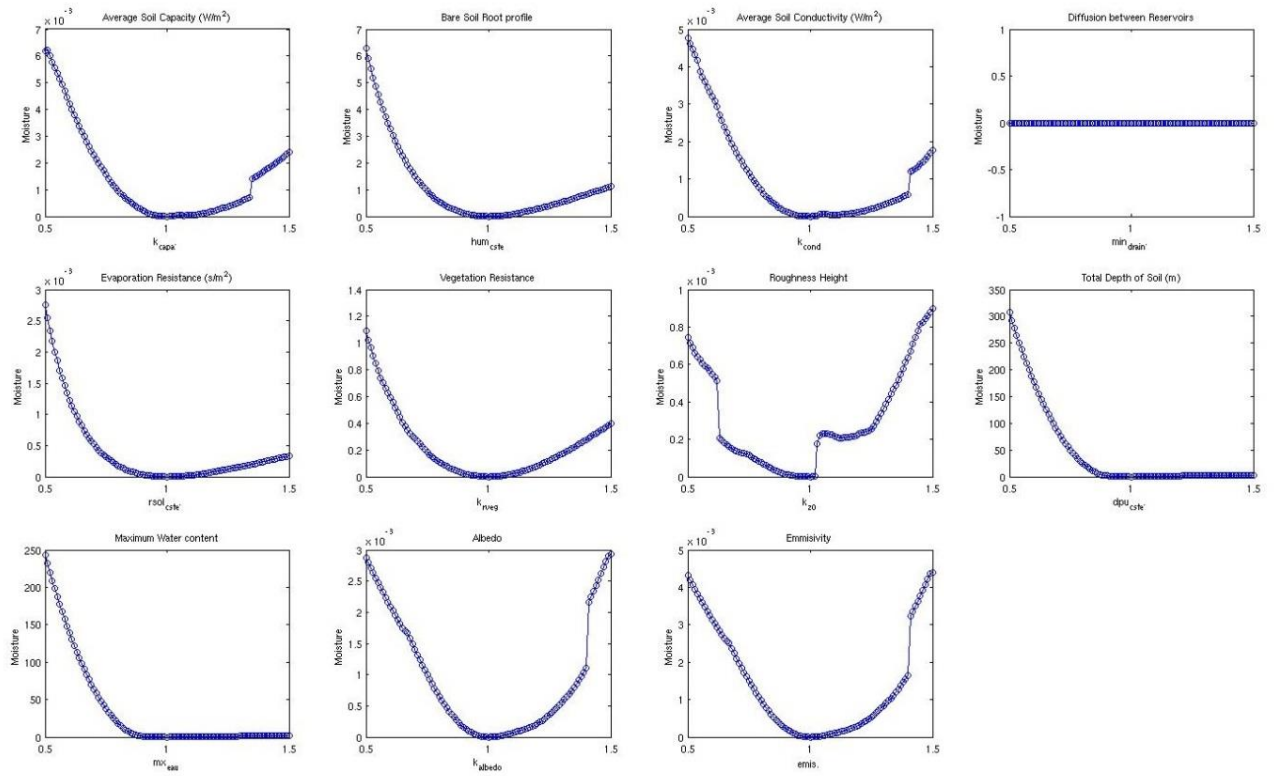


Figure C.2. Cost Function sensitivity analysis for land surface temperature using SMOSREX forcing, during a month, from 02-06-2006.

In Figure C.2, SMOSREX forcing cost function sampling is presented. The cumulated square difference of simulated and observed temperature is presented. Sampling was made during a month, starting on June the 2th, 2003, with grassland. The sampling gives us an idea of the influence of the variation of a single parameter in model variables.

The parameter hierarchy for soil temperature cost function is: k_{z0} , dpu_{cste} , k_{rveg} , mx_{eau} , hum_{cste} , k_{capa} , k_{cond} , k_{emis} , k_{albedo} and $rsol_{cste}$. As it can be seen, hierarchy does not match the hierarchy computed with the gradient method. This is due to the forcing and the specific characteristics of the site. The shortwave solar radiation and the wind speed in this site are not so smooth, introducing perturbation in the computation of the soil temperature. Changing the parameters in the cost function sampling introduce new perturbation that are not accounted when computing temperature gradients. Indeed, the cost function method perturbs parameters from -50% to +50% their prior value, opposite to the gradient method, when the parameters are perturbed only in 1% their prior value and thus the gradients represent accurately the exchange rate between the temperature for each unit of parameter.